MATHEMATICS OF OPERATIONS RESEARCH
Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 0364-765X | EISSN 1526-5471 | 00 | 0000 | 0001

INFORMS
DOI 10.1287/xxxx.0000.0000
© 0000 INFORMS

Chasing Demand: Learning and Earning in a Changing Environment

N. Bora Keskin
Fuqua School of Business, Duke University, Durham, NC 27708, bora.keskin@duke.edu,

Assaf Zeevi
Graduate School of Business, Columbia University, New York, NY 10027, assaf@gsb.columbia.edu,

We consider a dynamic pricing problem in which a seller faces an unknown demand model that can change over time. The amount of change over a time horizon of T periods is measured using a variation metric that allows for a broad spectrum of temporal behavior. Given a finite variation "budget," we first derive a lower bound on the expected performance gap between any pricing policy and a clairvoyant who knows a priori the temporal evolution of the underlying demand model, and then design families of near-optimal pricing policies, the revenue performance of which asymptotically matches said lower bound. We also show that the seller can achieve a substantially better revenue performance in demand environments that change in "bursts" than in demand environments that change "smoothly," among other things quantifying the net effect of the "volatility" in the demand environment on the seller's revenue performance.

Key words: Revenue management, pricing, model uncertainty, changing environment, sequential estimation, exploration-exploitation, regret.
MSC2000 subject classification: Primary: 93E35; secondary: 62L12, 60G35, 90B50
OR/MS subject classification: Primary: Probability, stochastic model applications; secondary: statistics, estimation
History: Received December 25, 2013; revised May 15, 2015, October 19, 2015, and April 22, 2016.

1. Introduction.

1.1. Background and overview. Pricing under demand uncertainty often involves a tradeoff between *learning* about customers' sensitivity to price variations, and *earning* short-term revenues. As a motivating example, consider the practice of evaluating loan applications in the financial sector. Because the applications are evaluated and approved on an individual basis, commercial banks and other financial institutions that sell short-term loans can offer a different interest rate to every customer. As noted by Phillips [26] this particular transaction structure in consumer lending, which is called *customized pricing*, offers relatively seamless opportunities for price experimentation to learn about customer behavior. Representing price sensitivity of customers in the form of a demand curve, a firm can use historical as well as real-time sales data to form estimates of the latter, while concurrently accumulating revenues from new sales. A key question in this context concerns the "perishability" of useful sales data, primarily due to changes in the demand environment.

Studies on dynamic pricing with demand model uncertainty have, by and large, focused almost exclusively on stylized settings where the demand environment, which is to be explored, does not change over time. The main focus of this paper is to extend this literature by formulating and studying a *time-varying* demand environment, and identifying some qualitative insights that arise from the learning-and-earning tradeoff in that case.

The particular learning-and-earning problem we consider has the following key features: (a) there is a seller who can dynamically change the price of its product over time; (b) the seller can observe the demand for its product, which depends on price and some unknown demand parameters; and (c) the unknown demand parameters can change over time. The seller's goal is to accumulate maximal revenues over a given time horizon, which could be achieved either by focusing on immediate revenues, or by learning the demand parameters to increase future revenues, or some combination thereof. Problem feature (c), which is the novel element in this study, motivates the seller to keep track of changing market conditions. We quantify the total amount of change over the time horizon using a variation metric in the demand model parameters, and measure the performance of a dynamic pricing policy using the growth rate of its *regret*: the expected revenue loss of a policy, as a function of the time horizon T, compared to a clairvoyant that knows the changing demand parameters. As will be explained in detail later, we first derive a lower bound on the minimum achievable growth rate of the regret, which *must* be incurred by *any* admissible policy, and then construct policies which admit a matching upper bound, and are hence optimal in order.

1.2. Main contributions and qualitative insights.

Summary of high level contributions. This paper makes three main contributions to the literature on dynamic pricing with demand model uncertainty. One main insight highlights a sharp difference that exists between "smooth" and "bursty" changes in a demand environment. This is manifest both in terms of the best achievable revenue performance as well as in the markedly different structure of (asymptotically) optimal policies in each case. Second, our analysis addresses a three-way tradeoff between learning, earning, and *information depreciation*; the former two are clearly central in dynamic pricing problems when there is uncertainty about the model, while the latter is driven by the fact that uncertainty itself is shifting over time. To see the role of information depreciation in this tradeoff, we derive several new results that characterize the complexity of the problem. Comparing our results with previous performance bounds obtained in recent studies on dynamic pricing with demand learning, we are able to precisely quantify the net effect of a changing demand environment on the seller's aggregate revenue. Finally, the policies we construct provide simple yet interesting guidelines for experimentation in changing environments. In particular, in the case of smooth changes we develop a weighted least squares estimation procedure that discounts older observations at an (asymptotically) optimal rate, whereas in the case of bursty changes we build a joint pricing and detection policy that repetitively tests if there has been a significant change in the environment.

On smooth versus bursty changes. In this paper we identify two families of changing demand environments that stand in stark contrast in terms of (a) the best achievable revenue performance, and (b) the use of pricing as a learning tool. The first family of demand environments is characterized by *smooth* changes (see the setting formulated in Section 2 and studied in Section 3), whereas the second family of demand environments is characterized by *bursty* changes (see Section 4). With regard to (a), the case of bursty changes seems to present a harsher environment at first glance, simply because at any given time the accumulated demand information can become worthless due to an abrupt change in the demand model. Somewhat surprisingly, our analysis proves the opposite. The essential intuition behind this observation is that gradual changes can practically be undetectable, and lead to substantial revenue loss in the case of smooth changes (see the proof and discussion of Theorem 1). With regard to (b), our analysis offers distinct ways to implement successful price experimentation in the two families of changing environments mentioned above. Knowing that undetected changes will lead to severe inaccuracies in estimation, the seller needs to discount the weight of older demand observations while estimating the demand curve in smoothly evolving environments. This gives rise to two practical price experimentation policies: moving windows; and decaying weights (akin to exponential smoothing). These maintain a near-optimal balance between learning, earning, and "information depreciation" (see the proof and discussion of Theorem 2). In the case of bursty changes, we construct a novel detection policy that can simultaneously detect and learn changes, incurring significantly smaller regret than the one characterizing smooth changes (see the proof and discussion of Theorem 4).

Information depreciation in changing environments. A distinguishing feature of our analysis is the explicit tradeoff between learning a demand curve, earning immediate revenues, and weighing down obsolete sales data. While the dual tradeoff between learning and earning has been studied extensively in the literature, there is limited work on the three-way tradeoff between learning, earning, and information depreciation. A key question here is whether a seller in a *changing* environment should collect information faster (or slower) than a seller in a *static* environment. More rapid information collection is desirable because the seller needs to constantly adapt to time-varying market conditions. On the other hand, slower information collection might also seem preferable because any piece of collected information will lose its value over time, implying that excessive attempts to accumulate information can cost more than its marginal value. The answer to these tradeoffs will depend on how "information" is defined. In static demand environments, the definition of information is fairly obvious; see, e.g., Keskin and Zeevi [22]. In changing environments, information collected up to a certain point in time only represents a *nominal* amount because part of this information becomes obsolete over time. It turns out that this notion can be quantified by considering the smallest eigenvalue of a suitably weighted Fisher information matrix, which measures the *relevant* amount of information in period t (see Section 3.2). Based on these definitions, one can revisit the tradeoffs related to the rate of information collection: a seller facing temporal demand changes should collect a larger amount of nominal information, but maintain a smaller amount of relevant information than a seller facing no demand change. The gap between the nominal and relevant information describes the near-optimal *information-depreciation rate* in the various demand settings studied in this paper, and moreover, enables us to quantify the time value of information in changing demand environments. For example, if we consider a policy that recalls only the data observed within a moving window, the ratio of the window size to the time horizon describes how fast the policy depreciates information. In light of this, we introduce an *information depreciation factor* defined as the ratio of the near-optimal moving window size in a given environment to time horizon T. On the extreme end of the spectrum, in a static environment, the information depreciation factor is equal to 1. This paper identifies the value of the information depreciation factors in some non-stationary settings, and hence quantifies the extent of information depreciation.

Organization of the paper. This section ends with a review of relevant literature. In Section 2 we formulate the problem, and in Section 3 we analyze it by first deriving a lower bound on the revenue loss of any given policy, and then designing near-optimal policies that achieve the loss rate in said lower bound. In Section 4 we consider demand environments that change in bursts, and construct a near-optimal policy whose performance is substantially better than the near-optimal performance observed in the case of smooth changes. Section 5 extends the results in Section 3 to the case of more rapidly changing demand environments, presenting a range of results that characterize the impact of the volatility in demand environment on the revenue performance. Section 6 contains some concluding remarks and a numerical example demonstrating the performance of our policies. Proofs of all results are in appendices, though proof sketches communicating key intuitive ideas are detailed in the main body.

1.3. Related literature.

In recent years, the tradeoff between learning and earning has become a prominent area of study in the literature on dynamic pricing and revenue management [see, e.g., 25, 1, 7, 16, 18, 10, 14, 22, 31], as well as in the broader operations management context [see, e.g., 19]. However, the vast majority of the studies in this area focus on learning in static environments in the sense that the ambient problem setting is unknown but does not change over time. One of the goals of the present paper is to provide a fairly general treatment of learning and earning in dynamically evolving environments, to study its implications on the value of price experimentation, and to illustrate the design of dynamic pricing policies that perform well in such settings.

In the economics literature, there has been considerable effort towards characterizing optimal learning policies in the presence of Markovian shifts in the demand model. As part of that effort, Balvers and Cosimano [3] and Beck and Wieland [4] examine dynamic control problems with autoregressive changes in underlying market-response model, while Rustichini and Wolinsky [28] and Keller and Rady [21] focus on similar problems in which underlying demand parameters evolve according to a two-state Markov chain. In all of these studies, the decision maker is assumed to know the transition rule for the time-varying (and unknown) parameters.

Another research stream that targets tracking problems is the statistics literature on change-point detection. As discussed in the survey papers by Lai [23] and Shiryaev [29], the essential motivation for change-point detection has been military and quality control applications, making these problems distinct from the tracking problems addressed in this study: in traditional change-point detection, the uncertainty is essentially about the time of change, and it is assumed that the decision maker knows exactly which model structure will be in place before and after the change; and this literature typically considers only a "passive" observation process, namely, the decision maker cannot influence the measurements being taken. These assumptions do not hold in the dynamic pricing applications that motivate our work. In the broader context of abruptly changing environments, Garivier and Moulines [17] study the performance of moving window policies in a multiarmed bandit problem. As mentioned in Section 1.2, and further discussed in the subsequent sections, the analysis in our formulation prescribes the use of "smooth" information-depreciation policies (such as moving windows and decaying weights) for gradually changing environments, and detection policies for abruptly changing environments (see Sections 3 and 4 for details).

In the operations research and management science (OR/MS) literature, the most notable studies on tracking problems are those of Aviv and Pazgal [2], Besbes and Zeevi [8], Chen and Farias [12] and den Boer [13]. Aviv and Pazgal [2] consider a revenue management problem with finite initial inventory, and construct a near-optimal policy, assuming that the underlying demand environment evolves according to a discrete-state-space Markov chain, the transition structure of which is known to the seller. Besbes and Zeevi [8] consider a dynamic pricing problem in which the demand environment changes at an unknown time in the sales horizon. Assuming that the seller has perfect knowledge about the demand curves before and after the change, they characterize an asymptotically optimal policy for jointly pricing and learning said change. More recently, Chen and Farias [12] find near-optimal policies in a dynamic pricing problem such that the market size evolves in a Markovian fashion whereas the price-sensitivity of customers remains stationary, and den Boer [13] studies well-performing pricing policies in a similar problem in which the market size is unknown and can change over time, but the price-sensitivity of customers is known with certainty. In contrast to the economics, statistics, and OR/MS studies mentioned above, we consider a fairly general formulation in which (i) the unknown demand model faced by the seller can change over time, resulting in unobservable changes in the market size *and* the price-sensitivity of consumers, and (ii) the decision maker does not know the particular dynamics of changes in the demand model, and our modeling paradigm affords a broad spectrum of such changes.

Our work is partially related to time series forecasting methods such as moving average and exponential smoothing, first formulated by Brown [11], later analyzed by Holt [20] and Winters [32], and studied extensively since. The main goal of such forecasting methods is to predict future values of a time series, the past values of which are observed in a noisy environment. A distinctive feature of our work is that the time series data in our formulation, which is composed of demand quantities, depends on the seller's pricing policy, and this necessitates the implementation of price experimentation to facilitate active learning. Therefore, "learning" and "information depreciation" need to be carried out simultaneously in our dynamic pricing problem. More importantly, assuming that the particular dynamic structure of the underlying price-response curve is unknown, we show that the seller can achieve an asymptotically optimal balance between learning and information depreciation by discounting older information using a *polynomial* weighting scheme.

At a high level, our work is also related to the literature on sequential stochastic optimization problems, which are often tackled by means of stochastic approximation type methods. In theory, it is possible to use stochastic approximation in dynamic pricing contexts, but unlike our work the vast majority of the stochastic approximation literature has focused on static environments. In parallel to our work, Besbes et al. [6] have examined how methods in the stochastic approximation literature can be used in changing environments, and there are high level similarities between the changing environments in our paper and in Besbes et al. [6] because both papers use a notion of cumulative variation in the underlying response functions. But, our work can be distinguished in several dimensions. Stochastic approximation type methods are designed to locally estimate the gradient of an unknown objective function and prescribe moving in the estimated gradient direction using a particular sequence of step sizes. In contrast, the least squares estimation methods we employ provide global estimates of the objective function and do not rely on predetermined step size sequences; hence, our policies are structurally different from those in Besbes et al. [6]. More importantly, we construct and analyze a variety of practically implementable tracking policies such as (1) moving windows, (2) decaying weights, and (3) joint detection-and-estimation. In contrast, the work by Besbes et al. [6] is based on the idea of repetitively restarting a stochastic gradient policy. As pointed out by Besbes et al. [6], this approach is not intended for implementation purposes but is rather meant to provide a "proof of concept." In particular, the gradient descent approach is in general not very suitable for pricing problems as this policy relies on local estimates and may frequently change prices in a highly unstructured way. In Section 6.3, we present a numerical experiment comparing the performance of our policies and stochastic approximation policies. In addition to these differences, our work differs from that of Besbes et al. [6] in terms of our identification of the sharp contrast between smooth and bursty changes, and the assumptions it makes on the decision maker's knowledge insofar as the cumulative variation in the environment. Once our formulation has been laid out, and the main results have been communicated, we will resume a more detailed discussion of these distinctions in Section 6.2.

2. Problem formulation.

Basic model elements. Consider a firm, called *the seller*, that sells a product over a time horizon of T periods. In each period $t = 1, 2, \ldots$ the seller chooses a price p_t for its product from a given interval $[\ell, u]$, where $0 < \ell < u < \infty$. After setting the price p_t, the seller observes demand D_t, which is given by

$$D_t = \alpha_t + \beta_t p_t + \epsilon_t \qquad \text{for } t = 1, 2, \ldots \tag{2.1}$$

where $\alpha_t \in \mathbb{R}$, $\beta_t \in \mathbb{R}_-$ are the demand model parameters, which are unknown to the seller, and ϵ_t are unobservable demand shocks. Assume that $\{\epsilon_t\}$ are independent and identically distributed random variables with mean zero and variance σ^2, and that there exists a positive constant x_0 such that $\mathbb{E}[\exp(x\epsilon_t)] < \infty$ for all $|x| \le x_0$ and all t. An important example is where $\epsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, but it is perhaps useful to note that the homogeneity assumption is not essential; it suffices that the variance of ϵ_t is bounded, and that the exponential moment condition holds uniformly. For notational brevity, we let $\theta_t = (\alpha_t, \beta_t)$ denote the vector of unknown demand parameters in period t, and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots)$ denote the sequence of demand parameter vectors. Let Θ be a compact rectangle in $\mathbb{R}\times\mathbb{R}_-$, from which the values of θ_t are chosen. Given a parameter vector $\theta=(\alpha,\beta)\in\Theta$ and a price $p\in[\ell,u]$, the seller's expected single-period revenue function is

$$r(p,\theta):=p\,(\alpha+\beta p). \tag{2.2}$$

We denote by $\varphi(\theta)$ the feasible price that maximizes the function $r(\cdot,\theta)$, that is

$$\varphi(\theta):=\arg\max\{r(p,\theta):\, p\in[\ell,u]\}. \tag{2.3}$$

To ensure that the revenue-maximizing price is always feasible, we assume that $\varphi(\theta)$ lies in the interior of $[\ell,u]$ for all $\theta=(\alpha,\beta)\in\Theta$. Thus, β is strictly negative and $\varphi\big((\alpha,\beta)\big)=-\alpha/(2\beta)$ for all $(\alpha,\beta)\in\Theta$.

Changing demand environment: the constant-budget problem. We measure the amount of change in T periods with the following quadratic variation metric: define a partition of $\{1,\ldots,T\}$ as any set of periods $\{t_0,t_1,\ldots,t_K\}$ satisfying $1\le t_0<\ldots<t_K\le T$ for some $K=1,2,\ldots,$ and denote by $\mathcal{P}$ the set of all partitions of $\{1,\ldots,T\}$. Given time horizon T, and a demand vector sequence $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_T)$, let

$$V_{\boldsymbol{\theta}}(T)\;:=\;\sup_{\{t_0,t_1,\ldots,t_K\}\in\mathcal{P},K\ge1}\left\{\sum_{k=1}^{K}\|\theta_{t_k}-\theta_{t_{k-1}}\|^2\right\}, \tag{2.4}$$

where $\|\cdot\|$ denotes the Euclidean norm of a vector. The values of θ_t are chosen from Θ such that the demand vector sequence $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_T)\in\Theta^T$ satisfies

$$V_{\boldsymbol{\theta}}(T)\le B\qquad\text{for } T=1,2,\ldots \tag{2.5}$$

where $B>0$. For notational brevity, we denote the set of demand vector sequences $\boldsymbol{\theta}$ satisfying (2.5) as follows:

$$\mathcal{V}(T,B)=\{\boldsymbol{\theta}:V_{\boldsymbol{\theta}}(T)\le B\}. \tag{2.6}$$

Inequality (2.5) describes a setting in which nature has a finite quadratic variation budget to change the demand parameters throughout the time horizon. We refer to this setting as the *constant-budget problem*. In Section 4, we analyze a special case of the constant-budget problem in which the changes occur in bursts, and in Section 5, we extend our results to the case of more rapidly changing environments where the upper bound in (2.5) can depend on (and increase with) T.

There is a natural upper bound on the quadratic variation metric: $V_{\boldsymbol{\theta}}(T)\le\lambda T$ for all $\boldsymbol{\theta}\in\Theta^T$, where $\lambda=\max\big\{\|\theta-\tilde\theta\|^2:\theta,\tilde\theta\in\Theta\big\}$. Therefore, we assume without loss of generality that $B\le\lambda T$. (Note that, if $B>\lambda T$, then we repeat our entire analysis by replacing B with $\tilde B=\lambda T$ in (2.5), and obtain the same performance guarantees because $\tilde B<B$.)

Pricing policies, induced probabilities, and performance metric. Let H_t denote the vectorized history of demands and prices observed through the end of period t, that is, $H_t=(D_1,p_1,\ldots,D_t,p_t)$. Define a *policy* as a sequence of functions $\pi=(\pi_1,\pi_2,\ldots)$, where π_1 is a constant function, and for all $t=1,2,\ldots,$ π_{t+1} is a function from $\mathbb{R}^{2t}$ into $[\ell,u]$, mapping H_t to the price that will be charged in period $t+1$. Any such policy π constructs a nonanticipating price sequence $p=(p_1,p_2,\ldots)$, where p_t is determined by the function π_t, and hence adapted to H_{t-1}. Note that this definition excludes randomized pricing policies.

Given a sequence of demand parameter vectors $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots)$ and a pricing policy π, we define a family of probability measures on the sample space of demand sequences $D=(D_1,D_2,\ldots)$ as follows. Let $\mathbb{P}^\pi_{\boldsymbol{\theta}}$ be a probability measure satisfying

$$\mathbb{P}^\pi_{\boldsymbol{\theta}}(D_1\in d\xi_1,\ldots,D_T\in d\xi_T)\;=\;\prod_{t=1}^{T}\mathbb{P}_\epsilon(\alpha_t+\beta_t p_t+\epsilon_t\in d\xi_t)\qquad\text{for }\xi_1,\ldots,\xi_T\in\mathbb{R}, \tag{2.7}$$

where $\mathbb{P}_\epsilon(\cdot)$ is the probability measure governing the random variables $\{\epsilon_t\}$, and $p=(p_1,p_2,\ldots)$ is the price sequence formed under policy π and demand realization $D=(D_1,D_2,\ldots)$.

The performance metric we use in this paper is T-period *regret*, defined as

$$\mathcal{R}^\pi(T,B) \;=\; \sup\bigl\{\Delta^\pi_{\boldsymbol\theta}(T) : \boldsymbol\theta\in\mathcal{V}(T,B)\bigr\}, \tag{2.8}$$

where for $T=1,2,\ldots$ and $\boldsymbol\theta=(\theta_1,\ldots,\theta_T)$

$$\Delta^\pi_{\boldsymbol\theta}(T) \;=\; \mathbb{E}^\pi_{\boldsymbol\theta}\biggl\{\sum_{t=1}^{T}\biggl(1-\frac{r(p_t,\theta_t)}{r^*(\theta_t)}\biggr)\biggr\}, \tag{2.9}$$

$\mathbb{E}^\pi_{\boldsymbol\theta}(\cdot)$ is the expectation operator associated with the probability measure $\mathbb{P}^\pi_{\boldsymbol\theta}(\cdot)$, $r^*(\theta):=r\bigl(\varphi(\theta),\theta\bigr)$ is the optimal single-period revenue function, and $\mathcal{V}(T,B)$ is as defined in (2.6). The regret of a policy is the *worst-case* expected normalized revenue loss relative to a clairvoyant policy that knows the value of θ_t in every period. Given the normalization in (2.9), this can also be interpreted as the expected number of lost sales opportunities due to not knowing the underlying demand model. Under either interpretation, when the regret of a policy is sublinear in T, the policy is long-run-average optimal, and more generally, smaller regret corresponds to uniformly better revenue performance.

3. Analysis of the constant-budget problem.

3.1. A lower bound on regret. Our first result is a theoretical lower bound on the minimum achievable regret of any pricing policy in the constant-budget problem setting described in the preceding section.

Theorem 1. (lower bound on regret) *There exists a finite positive constant c such that $\mathcal{R}^\pi(T,B)\ge c\,B^{1/3}T^{2/3}$ for any pricing policy π and time horizon $T\ge 3$.*

Theorem 1 shows that the T-period regret of any given policy is at least on the order of $T^{2/3}$. A policy π that achieves the loss rate in Theorem 1, i.e., any policy π such that $\mathcal{R}^\pi(T)=O(T^{2/3})$ will hereafter be called *first-order optimal*, and *rate optimal* if the dependence on *both* B and T matches the lower bound.

Rough proof sketch. The main intuition behind this result is that nature can change the demand parameters in a gradual manner such that it is very costly to detect changes and learn the new demand curve after a change. By carefully choosing a parameter change with squared norm of order $T^{-1/3}$, nature makes sure that: either (i) no detection test can identify this change without incurring a loss of order $T^{1/3}$, or (ii) the cost of learning the new parameter vector is of order $T^{1/3}$. Within its change budget, nature can use $T^{1/3}$ such parameter changes, implying that any given policy must have a loss of order $T^{2/3}$, even if it is designed to simultaneously detect and learn. To prove arguments (i) and (ii), we use the Kullback-Leibler divergence to quantify the difference between the likelihood of events under the probability measures before and after a potential change. For a given policy, if the Kullback-Leibler divergence is smaller than a threshold η, then we derive argument (i) via hypothesis testing results in information theory. In particular, a Fano-type lower bound on the error probabilities in a detection problem [cf. 30, Theorem 2.2] implies that there is a significant probability of not detecting the potential change, which consequently leads to a revenue loss of order $T^{-1/3}N$ in the N periods following the potential change. On the other hand, if the Kullback-Leibler divergence is larger than the threshold η, then we note that, despite the small amount of change, the cost of gathering information on the new demand parameters is bounded away from zero. This implies that the revenue loss until the next change will be of order $T^{1/3}$, as expressed in argument (ii). If there are N periods between two changes, then arguments (i) and

(ii) imply that the revenue loss between these two changes is at least of order $(T^{-1/3}N) \wedge T^{1/3}$, where $\wedge$ denotes the minimum of two numbers. We therefore deduce that nature can cause a loss of order $T^{2/3}$ within T periods, by choosing N to be of order $T^{2/3}$ and spreading out potential changes throughout the time horizon. Q.E.D.

Discussion and key insights. The derivation of Theorem 1 brings forth a key insight about the type of policies that could perform well in the setting described in the preceding section. Specifically, the seller can face a sequence of smooth changes that are virtually undetectable, making any effort to detect changes perform poorly. Therefore, successful policies in this environment should not focus on detecting every single change, but instead, they need to *depreciate information* at some rate, with the hope that the negative effects of undetectable smooth changes will be filtered out sufficiently fast. The next subsection provides a general estimation procedure that implements this idea by assigning non-increasing weights to older demand observations.

Besbes et al. [6] provide a lower bound on regret that grows proportional to $T^{2/3}$ in a stochastic optimization setting, and their proof relies on the similar use of Kullback-Leibler divergence and Tsybakov's Theorem 2.2 [30], which is a commonly used proof technique in deriving such lower bounds [see also 7, 10]. Our lower bound in Theorem 1 establishes that the complexity of our parametric dynamic pricing problem is in the same order of magnitude as nonparametric stochastic optimization problems in changing environments.

3.2. A weighted least squares estimator. In what follows we describe a procedure to estimate θ_{t+1} given the history of demands and prices through the end of period t. Let $w^t = (w_1^t, \dots, w_t^t)$ be a $t \times 1$ vector of nonnegative real numbers. Given history vector H_t and weight vector w^t, set the *weighted least squares* estimator of θ_{t+1} to be

$$\hat{\theta}_{t+1} = \arg\min_\theta \{SSE_t(\theta, w^t)\}, \tag{3.1}$$

where $SSE_t(\theta, w^t) = \sum_{s=1}^t w_s^t (D_s - \alpha - \beta p_s)^2$ for $\theta = (\alpha, \beta)$. If the matrix $\left[\begin{smallmatrix} \sum_{s=1}^t w_s^t & \sum_{s=1}^t w_s^t p_s \\ \sum_{s=1}^t w_s^t p_s & \sum_{s=1}^t w_s^t p_s^2 \end{smallmatrix}\right]$ is invertible, then the solution of the weighted least squares problem (3.1) is

$$\hat{\theta}_{t+1} = \begin{bmatrix} \hat{\alpha}_{t+1} \\ \hat{\beta}_{t+1} \end{bmatrix} = \begin{bmatrix} \sum_{s=1}^t w_s^t & \sum_{s=1}^t w_s^t p_s \\ \sum_{s=1}^t w_s^t p_s & \sum_{s=1}^t w_s^t p_s^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{s=1}^t w_s^t D_s \\ \sum_{s=1}^t w_s^t D_s p_s \end{bmatrix}. \tag{3.2}$$

Let us now re-express (2.1) in the following compact form:

$$D_t = X_t \cdot \theta_t + \epsilon_t \qquad \text{for } t = 1, 2, \dots \tag{3.3}$$

where $X_t = \left[\begin{smallmatrix} 1 \\ p_t \end{smallmatrix}\right]$. Then, (3.2) and (3.3) imply that

$$\begin{aligned} \hat{\theta}_{t+1} - \theta_{t+1} &= \left(\sum_{s=1}^t w_s^t X_s X_s^\mathsf{T} \right)^{-1} \left(\sum_{s=1}^t w_s^t X_s X_s^\mathsf{T} (\theta_s - \theta_{t+1}) + \sum_{s=1}^t w_s^t X_s \epsilon_s \right) \\ &= \left(\mathcal{J}_t^t\right)^{-1} \mathcal{W}_t^t + \left(\mathcal{J}_t^t\right)^{-1} \mathcal{M}_t^t \quad \text{for all } t = 2, 3, \dots, \end{aligned} \tag{3.4}$$

where $\mathcal{J}_q^t$ is the empirical Fisher information given by

$$\mathcal{J}_q^t = \sum_{s=1}^q w_s^t X_s X_s^\mathsf{T}, \tag{3.5}$$

$\mathcal{W}_q^t = \sum_{s=1}^q w_s^t X_s X_s^\mathsf{T} (\theta_s - \theta_{t+1})$, and $\mathcal{M}_q^t = \sum_{s=1}^q w_s^t X_s \epsilon_s$. The first term on the right hand side of (3.4) is the *estimation inaccuracy* due to the changing environment, and the second term is the *estimation error* due to noise.

3.3. Unknown variation budget: designing first-order optimal policies. In this subsection we construct policies for the case where the seller does not know the variation budget B at the outset. We first define two families of pricing policies that rely on (i) price experimentation with a carefully chosen frequency, and (ii) the use of weighted least squares estimation with a particular sequence of weights. Then, we prove that these families of policies achieve $O(T^{2/3})$ regret in T periods, and hence are first-order optimal.

Price experiments. The policies we consider in this section conduct price tests with a certain frequency in the following manner: let $\kappa \geq 2$, and x_1, x_2 be two distinct test prices in $[\ell, u]$. To construct the set of periods, $\mathcal{X}_1$, $\mathcal{X}_2$, at which the test prices will be charged, let $n := \lceil \kappa T^{1/3} \rceil$, and

$$\mathcal{X}_i \ := \ \big\{ t = kn + i \ : \ k = 0, 1, 2, \ldots, \lfloor T/n \rfloor \big\} \qquad \text{for } i = 1, 2. \tag{3.6}$$

In period t, charge the price

$$p_t = \begin{cases} x_1 & \text{if } t \in \mathcal{X}_1 \\ x_2 & \text{if } t \in \mathcal{X}_2 \\ \varphi(\vartheta_t) & \text{otherwise,} \end{cases} \tag{3.7}$$

where ϑ_t is the truncated estimate that satisfies $\vartheta_t := \arg\min_{\vartheta \in \Theta} \{ \| \vartheta - \hat{\theta}_t \| \}$. In the above experimentation scheme, the frequency of price tests is $2/n$, which is of order $T^{-1/3}$.

Moving windows and gradually decaying weights. Our first policy estimates the unknown demand vector using only the most recent price tests within a moving time window, forgetting all data outside said window. The *moving window policy* with parameters κ, x_1, x_2, denoted by $M(\kappa, x_1, x_2)$, chooses prices according to (3.6) and (3.7), and uses a sequence of weight vectors $\{w^1, w^2, \ldots\}$ such that $w^t = (w_1^t, \ldots, w_t^t)$ for $t = 1, \ldots, T$, where

$$w_s^t = \begin{cases} 1 \text{ if } s \in \mathcal{X} \ \text{ and } \ s \geq t - n^2 \\ 0 \text{ otherwise} \end{cases} \tag{3.8}$$

for $1 \leq s \leq t$, and $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$.

Our second policy puts decreasing weights on older observations in a gradually decaying manner. The *decaying weights policy* with parameters μ, κ, x_1, x_2, denoted by $W(\mu, \kappa, x_1, x_2)$, selects prices according to (3.6) and (3.7), and uses a sequence of weight vectors $\{w^1, w^2, \ldots\}$ such that $w^t = (w_1^t, \ldots, w_t^t)$ for $t = 1, \ldots, T$, where

$$w_s^t = \begin{cases} \left(1 - \dfrac{t - \tilde{s}}{n^2} + \dfrac{(t - \tilde{s})^{1-\mu}}{n^2} \right)_+^{\frac{1}{\mu}} & \text{if } s \in \mathcal{X} \\ 0 & \text{otherwise} \end{cases} \tag{3.9}$$

for $1 \leq s \leq t$, $0 < \mu \leq 1$, and $\tilde{s} = s + \mathbb{I}\{s \in \mathcal{X}_1\}$. Under $W(\mu, \kappa, x_1, x_2)$, the weight given to any observation decreases smoothly via the decay parameter μ. An extreme choice for μ is 1, in which case the weights in (3.9) become $w_s^t = \big(1 - (t - \tilde{s} + 1)/n^2\big)_+ \mathbb{I}\{s \in \mathcal{X}\}$ for $1 \leq s \leq t$, implying that weights decay linearly over time. As μ approaches zero, we achieve slower decay rates.

Note that, for $M(\kappa, x_1, x_2)$ and $W(\mu, \kappa, x_1, x_2)$, the empirical Fisher information in the estimation problem (3.1-3.2) has the following form:

$$\mathcal{J}_t^t \ = \ \mathfrak{X} \, \mathcal{I}_t^t \tag{3.10}$$

for all $t \geq n^2$, where

$$\mathfrak{X} = \begin{bmatrix} 2 & x_1 + x_2 \\ x_1 + x_2 & x_1^2 + x_2^2 \end{bmatrix}, \tag{3.11}$$

and $\mathcal{I}_q^t = \frac{1}{2} \sum_{s=1}^{q} w_s^t \geq 0$ represents the "relevance" in period t of the information in period q.

Performance of the moving window and decaying weights policies. We now show that the two policy families defined above are first-order optimal. In our first result, we derive upper bounds on the aggregate estimation inaccuracy due to changes in demand parameters.

Lemma 1. (upper bound on aggregate estimation inaccuracy) *There exists a finite positive constant c_1, such that under either $M(\kappa,x_1,x_2)$ or $W(\mu,\kappa,x_1,x_2)$*

$$\sum_{t=n^2}^{T-1} \left\| \left(\mathcal{J}_t^t\right)^{-1} \mathcal{W}_t^t \right\|^2 \le c_1 T^{2/3} \tag{3.12}$$

almost surely for all $T=1,2,\ldots$ and $\boldsymbol{\theta}\in\mathcal{V}(T,B)$.

To derive Lemma 1, we first note that a change after periods s contributes to the estimation inaccuracy in period t if and only if $w_s^t>0$, i.e., the demand observation in period s has positive weight in period t. Therefore, in a changing environment, the seller needs to: (i) avoid giving excessive weight to past observations to limit the contribution of a parameter change to the estimation inaccuracy; and at the same time, (ii) give non-negligible weight to past observations to accumulate information. More formally, (i) can be viewed as an *information-depreciation* condition that guarantees that the norm of $\mathcal{W}_t^t$ grows sufficiently slowly. On the other hand, condition (ii), which can be interpreted as a *learning* condition, guarantees that the eigenvalues of $\mathcal{J}_t^t$ grow sufficiently fast. Moving windows and decay weights are two distinct ways to resolve the tradeoff between these information-depreciation and learning conditions. To obtain the bound in (3.12) for $M(\kappa,x_1,x_2)$, the size of the moving window should be small enough to meet the information-depreciation condition (i), but also large enough to meet the learning condition (ii). Similarly, for $W(\mu,\kappa,x_1,x_2)$, the weight decay rate should be fast enough to satisfy (i), and simultaneously slow enough to satisfy (ii). Lemma 1 states that the careful selection of window sizes and decay rates in (3.8) and (3.9), respectively, leads to an $O(T^{2/3})$ aggregate estimation inaccuracy, which grows at the first-order optimal rate described in Theorem 1.

Our second result characterizes how estimation errors due to noise decay over time.

Lemma 2. (exponential decay of estimation error due to noise) *Let π be either $M(\kappa,x_1,x_2)$ or $W(\mu,\kappa,x_1,x_2)$. Then there exists a finite positive constant ρ such that*

$$\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\left\{ \left\| \left(\mathcal{J}_t^t\right)^{-1} \mathcal{M}_t^t \right\| > z,\ \mathcal{I}_t^t > \gamma \right\} \ \le\ 4\, e^{-\rho (z\wedge z^2)\gamma} \tag{3.13}$$

for all $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_T)$, $z>0$, $\gamma>0$, and $t\ge 2$.

Lemma 2 states that the tail probability of the estimation error $\left(\mathcal{J}_t^t\right)^{-1}\mathcal{M}_t^t$ decays exponentially, and the rate of this decay is determined by the amount of relevant information in period t, namely $\mathcal{I}_t^t=\frac{1}{2}\sum_{s=1}^t w_s^t\ge 0$. Using Lemmas 1 and 2 we obtain the following performance bound for moving window and decaying weights policies.

Theorem 2. (first-order optimality) *Let π be either $M(\kappa,x_1,x_2)$ or $W(\mu,\kappa,x_1,x_2)$. Then there exists a finite positive constant C such that $\mathcal{R}^{\pi}(T)\le C\,T^{2/3}$ for all $T\ge 3$.*

Remark 1. The constant C in the preceding theorem is linear in B.

Discussion. The preceding theorem establishes the first-order optimality of the policies constructed in this subsection. The main intuition behind this result is a careful balancing of three goals: (i) learning; (ii) earning; and (iii) information depreciation. The experimentation scheme in (3.6-3.7), and the weights in (3.8-3.9) ensure that the information metric $\mathcal{I}_t^t$ is proportional to $T^{1/3}$, and as shown in Lemma 1, this leads to a first-order optimal balance between learning and information depreciation. In Theorem 2, we show that the same experimentation scheme and choice of weights also achieve a first-order optimal balance between learning and earning: by maintaining the relevant amount of information in the order of $T^{1/3}$, the seller guarantees that the aggregate losses due to estimation inaccuracy and estimation error are $O(T^{2/3})$. To keep the relevant amount of information at that level while depreciating historical information at the rate given in Lemma 1, the seller conducts $O(T^{2/3})$ price tests, implying that the cost of experimentation is also of order $T^{2/3}$. These two relations between experimentation cost, estimation error, and estimation inaccuracy provide a fine balance between learning, earning, and information depreciation, from which we derive the first-order optimal performance bound in Theorem 2.

3.4. Known variation budget: designing rate optimal policies. We now consider the case where the seller knows the variation budget B, and construct a family of policies that are rate optimal.

Price experiments. To adapt to a known B parameter, consider the following modification of the price experimentation scheme in Section 3.3. Given κ, let $n := \lceil \kappa B^{-1/3} T^{1/3} \rceil$, and

$$\mathcal{X}_i \;:=\; \left\{ t = kn^2 + (i-1)n + q \;:\; k = 0, 1, \ldots, \lfloor T/n^2 \rfloor,\, q = 1, \ldots, n \right\} \tag{3.14}$$

for $i = 1, 2$. The price to be charged in period t is given by

$$p_t = \begin{cases} x_1 & \text{if } t \in \mathcal{X}_1 \\ x_2 & \text{if } t \in \mathcal{X}_2 \\ \varphi(\vartheta_t) & \text{otherwise,} \end{cases} \tag{3.15}$$

where x_1 and x_2 are two distinct test prices in $[\ell, u]$, and ϑ_t is the truncated estimate that satisfies $\vartheta_t := \arg\min_{\vartheta \in \Theta} \{ \|\vartheta - \hat{\theta}_t\| \}$. Because $B \le \lambda T$, we ensure that $n \ge 2$ by choosing $\kappa \ge 2\lambda^{1/3}$, where λ is a constant independent of B and T. This experimentation scheme, like its counterpart in the preceding subsection, conducts price tests with a frequency of $2/n$, which is of order $T^{-1/3}$.

Moving windows. Consider the following modification of the moving window policy defined in Section 3.3: given $B > 0$, suppose that the *moving window policy* with parameters κ, x_1, x_2, denoted by $M_B(\kappa, x_1, x_2)$, charges the prices in (3.14-3.15), and uses a sequence of weight vectors $\{w^1, w^2, \ldots\}$ such that $w^t = (w^t_1, \ldots, w^t_t)$ for $t = 1, \ldots, T$, where

$$w^t_s = \begin{cases} 1 \text{ if } s \in \mathcal{X} \;\text{ and }\; s \ge t - n^2 \\ 0 \text{ otherwise,} \end{cases} \tag{3.16}$$

for $1 \le s \le t$, and $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$. (We also note that, as in Section 3.3, it is possible to construct a counterpart of $M_B(\kappa, x_1, x_2)$ that depreciates information with decaying weights rather than moving windows.) Under $M_B(\kappa, x_1, x_2)$, the empirical Fisher information in the estimation problem (3.1-3.2) has the form given in (3.10). This allows us to prove counterparts of Lemmas 1 and 2 for $M_B(\kappa, x_1, x_2)$, and consequently derive the following rate optimal performance guarantee.

THEOREM 3. (RATE OPTIMALITY) *Let π be $M_B(\kappa, x_1, x_2)$. Then there exists a finite positive constant C such that $\mathcal{R}^\pi(T, B) \le C\, B^{1/3} T^{2/3}$ for all $T \ge 3$.*

Recall that, as shown in Theorem 2, the moving window and decaying weights policies described in Section 3.3 achieve $O(T^{2/3})$ regret *without* relying on the a priori knowledge of B. Because the constant C in Theorem 2 grows linearly in B, the difference between that constant and the $O(B^{1/3})$ constant in Theorem 3 helps us quantify the "price" of adapting to an unknown variation budget. This difference stems from the fact that, when B is unknown, the seller can act as if B equals a known constant, and this would make the aggregate squared inaccuracy, namely $\sum_t \|(\mathcal{J}^t_t)^{-1} \mathcal{W}^t_t\|^2$, grow proportional to the quadratic variation, which is bounded above by B.

Besbes et al. [6] focus on the case of a known variation budget, and use a stochastic approximation policy that is restarted with a predetermined frequency to obtain a performance guarantee similar to Theorem 3 [see 6, Section 5.2, EGS algorithm]. We will revisit this policy in Section 6.3 and compare its performance with the performance of our policies in Section 3.3 that do not rely on the knowledge of B.

4. Learning and detection of bursty changes. In this section, we consider the case of bursty changes that are characterized by a positive minimum change constraint: suppose that there exists a positive constant δ satisfying

$$d_{\boldsymbol{\theta}} \;:=\; \inf\big\{\|\theta_t-\theta_s\| : \theta_t\neq\theta_s,\, 1\le s<t\le T\big\} \;\ge\; \delta. \tag{4.1}$$

In contrast to gradual and potentially undetectable changes that can happen in the setting studied in the preceding section, condition (4.1) states that any pair of distinct values attained by the sequence $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots)$ are at least δ apart from each other. This implies that changes happen in bursts; that is, whenever the demand vector changes, its Euclidean norm has to change by at least δ. Combined with condition (2.5), this implies that there can be at most $\bar C=\lceil B/\delta^2\rceil$ changes. An extreme example in the family of admissible changing environments described by conditions (2.5) and (4.1) is the case of a single change-point (over the entire time horizon). In comparison with traditional change-point detection problems, the distinguishing feature of our problem is the need to learn the demand parameters before and after the change, which makes it more difficult to detect the change-point. Another example is the case of switching back and forth between two distinct values of demand parameters. The repetitive nature of this example requires conducting multiple detection tests, which could lead to multiple false alarms before a non-spurious change-point is detected, unlike single change-point detection tests. In general, the demand parameter sequence $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots)$ can take on $\bar C+1$ distinct values, all of which are initially unknown to the seller.

With the addition of the bursty change condition (4.1), we update our performance metric as follows: let

$$\mathcal{R}^{\pi}(T,B,\delta) \;=\; \sup\big\{\Delta^{\pi}_{\boldsymbol{\theta}}(T):\boldsymbol{\theta}\in\mathcal{V}(T,B,\delta)\big\}, \tag{4.2}$$

where $\Delta^{\pi}_{\boldsymbol{\theta}}(T)$ is as defined in (2.9), and $\mathcal{V}(T,B,\delta)=\{\boldsymbol{\theta}:V_{\boldsymbol{\theta}}(T)\le B, d_{\boldsymbol{\theta}}\ge\delta\}$. Note that $\mathcal{V}(T,B,\delta)$ is a subset of its counterpart in the preceding section, namely $\mathcal{V}(T,B)$, which is given in (2.6). One of the key questions we would like to investigate in this section is whether we can achieve significantly smaller regret by imposing the bursty change condition (4.1) on the set of admissible demand parameter sequences.

4.1. Dynamic pricing with simultaneous learning and detection. Assuming the seller knows that (4.1) holds, we design a well-performing pricing policy that detects change-points and learns unknown demand parameters simultaneously.

Price experiments. Let κ and η be two positive real numbers, and x_1, x_2 be two distinct test prices in $[\ell,u]$. The *detection policy* with parameters η,κ,x_1,x_2, denoted by $D(\eta,\kappa,x_1,x_2)$, divides the time horizon into cycles of $n:=\lceil\kappa T^{1/2}\rceil$ periods, and conducts price experiments in the first $2m$ periods of every cycle, where $m:=\lceil\kappa\log T\rceil$. To be precise, the sets of periods at which $D(\eta,\kappa,x_1,x_2)$ conducts price experiments are given by

$$\mathcal{X}_{ik} \;:=\; \big\{t=kn+(i-1)m+q \;:\; q=1,2,\ldots,m\big\} \tag{4.3}$$

for $i=1,2$, and $k=0,1,2,\ldots,\lfloor T/n\rfloor$. In period t, $D(\eta,\kappa,x_1,x_2)$ charges the price

$$p_t=\begin{cases} x_1 & \text{if } t\in\mathcal{X}_1\\ x_2 & \text{if } t\in\mathcal{X}_2\\ \varphi(\vartheta_t) & \text{otherwise,}\end{cases} \tag{4.4}$$

where $\mathcal{X}_i=\bigcup_k\mathcal{X}_{ik}$ for $i=1,2$, and ϑ_t is the truncated estimate that satisfies $\vartheta_t:=\arg\min_{\vartheta\in\Theta}\{\|\vartheta-\hat\theta_t\|\}$. In this experimentation scheme, the frequency of price tests is $2m/n$, which is of order $T^{-1/2}\log T$.

Joint change-point detection and parameter estimation. We will now describe a detection scheme that dynamically updates the weight vector sequence $\{w^t\}$ of the weighted least squares estimator in (3.2) by placing zero weight on periods that precede a detected change-point. Fix $k \in \{0, \ldots, \lfloor T/n \rfloor\}$. Denote by $\bar{D}_{ik}$ the average demand observed during the periods in $\mathcal{X}_{ik}$, that is

$$\bar{D}_{ik} \;:=\; m^{-1} \textstyle\sum_{t \in \mathcal{X}_{ik}} D_t. \tag{4.5}$$

Construct the binary-valued detection process $\chi := \{\chi_0, \chi_1, \ldots\}$ as follows: fix $\chi_0 = 1$, and define the *latest detection cycle* as $L(k) := \max\{\tau \le k : \chi_\tau = 1\}$. With this formalism a price experiment in period s occurs after the latest detection (prior to period t) if and only if $s > nL(t/n)$. For every cycle $k = 0, 1, \ldots, \lfloor T/n \rfloor$, let

$$\chi_{k+1} = \begin{cases} 1 \text{ if } \sup_{i,k'} \left\{ |\bar{D}_{ik} - \bar{D}_{ik'}| : i = 1, 2,\, L(k) \le k' < k \right\} > \eta \\ 0 \text{ otherwise,} \end{cases} \tag{4.6}$$

where the supremum of the empty set is taken to be $-\infty$. The detection test in (4.6) repeatedly checks whether there has been a change in the average demand observed since the latest detection. To make a comparison, it is necessary to compute at least one average demand estimate after each detection, and hence, it is not possible to have two consecutive detections: for any given cycle k with $\chi_k = 1$, we have $L(k) = k$ and there is no k' satisfying $L(k) \le k' < k$, implying that $\chi_{k+1} = 0$.

In cycle $k = 0, 1, 2, \ldots, \lfloor T/n \rfloor$, the seller observes $\{\bar{D}_{ik'}\}_{i=1,2,\, k'=0,1,\ldots,k}$ by the end of period $(k+1)n$, which implies that χ_{k+1} is a function that maps demand in the first $(k+1)n$ periods, $D_1, D_2, \ldots, D_{(k+1)n}$, into $\{0, 1\}$. Based on the realization of the detection process χ, $D(\eta, \kappa, x_1, x_2)$ uses the following weights for estimation:

$$w_s^t = \begin{cases} 1 \text{ if } s \in \mathcal{X} \text{ and } s > nL(t/n) \\ 0 \text{ otherwise,} \end{cases} \tag{4.7}$$

for $1 \le s \le t$, where $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$. In other words, $D(\eta, \kappa, x_1, x_2)$ recalls all available data as long as it does not detect a change, and forgets all past data immediately after it detects a change (hence it restarts learning whenever the value of the process χ switches from 0 to 1).

4.2. Performance of the detection policy. In this section, we prove that the regret of the detection policy described above is of order $T^{1/2} \log T$. To derive this result, we first define the random times at which detections happen. Using these random times, we decompose the regret according to different sources of loss, and then bound each of them.

Suppose that the unknown parameter sequence $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots)$ has $\mathcal{C}$ change-points in the first T periods, and denote by t_j^* the j^{th} change-point. That is, let $1 = t_0^* < t_1^* < \ldots < t_{\mathcal{C}}^* < t_{\mathcal{C}+1}^* = T + 1$, where $t_j^* = \inf\{t \ge t_{j-1}^* : \theta_t \ne \theta_{t_{j-1}^*}\}$ for $j = 1, 2, \ldots, \mathcal{C}$. Recalling that $D(\eta, \kappa, x_1, x_2)$ divides the time horizon into cycles of $n = \lceil \kappa T^{1/2} \rceil$ periods, we let $\tau_j^* := \lfloor (t_j^* - 1)/n \rfloor$ be the cycle of the j^{th} change-point, and $\hat{\tau}_j^+$ and $\hat{\tau}_j^-$ be the cycles containing the first and second declared detections, respectively, between the cycles of j^{th} and $(j+1)^{\text{st}}$ change-points (if there is no detection between the j^{th} and $(j+1)^{\text{st}}$ change-points then we set $\hat{\tau}_j^+ = \hat{\tau}_j^- = \tau_{j+1}^*$). More formally, define $\hat{\tau}_0^+ := 0$, and put

$$\hat{\tau}_j^+ \;:=\; \inf\{\tau > \tau_j^* : \chi_\tau = 1\} \wedge \tau_{j+1}^* \qquad \text{for } j = 1, 2, \ldots, \mathcal{C}, \tag{4.8}$$

$$\hat{\tau}_j^- \;:=\; \inf\{\tau > \hat{\tau}_j^+ : \chi_\tau = 1\} \wedge \tau_{j+1}^* \qquad \text{for } j = 0, 1, \ldots, \mathcal{C}, \tag{4.9}$$

where the infimum of the empty set is taken to be ∞. The definitions in (4.8-4.9) are interpreted as follows: in (4.8), we label the first declared detection after the j^{th} change-point as a “correct” detection, regardless of delay. As will be shown below, the loss due to delays of correct detections is reasonably small for our policy. If a correct detection is followed by any further such declarations before the $(j+1)^{\text{st}}$ change-point, we consider all these subsequent events as “false” detections because there is no actual underlying change. In (4.9), we label the earliest of the false detections after the j^{th} change-point.

In the context of joint change-point detection and parameter estimation, the loss of a policy stems from four sources: (i) delay in correct detections; (ii) false alarms; (iii) estimation errors due to noise; and (iv) cost of experimentation. Let us first decompose the T-period regret with respect to losses due to (i-iii) and (iv). Because the cardinality of the experimentation set $\mathcal{X}=\mathcal{X}_1\cup\mathcal{X}_2$ is at most $2m\lceil T/n\rceil\le 8T^{1/2}\log T$, we have

$$\begin{aligned}
\Delta^{\pi}_{\boldsymbol{\theta}}(T) &= \mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{\sum_{t=1}^{T}\left(1-\frac{r(p_t,\theta_t)}{r^*(\theta_t)}\right)\right\}\\
&= \mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{\sum_{t=1}^{T}\left(1-\frac{r(p_t,\theta_t)}{r^*(\theta_t)}\right)\mathbb{I}\{t\in\mathcal{X}\}\right\} + \mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{\sum_{t=1}^{T}\left(1-\frac{r(p_t,\theta_t)}{r^*(\theta_t)}\right)\mathbb{I}\{t\notin\mathcal{X}\}\right\}\\
&\le 2m\left\lceil\frac{T}{n}\right\rceil + \mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{\sum_{t=1}^{T}\left(1-\frac{r(p_t,\theta_t)}{r^*(\theta_t)}\right)\mathbb{I}\{t\notin\mathcal{X}\}\right\}\\
&\le 8T^{1/2}\log T + \mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{\sum_{t=1}^{T}\left(1-\frac{r(p_t,\theta_t)}{r^*(\theta_t)}\right)\mathbb{I}\{t\notin\mathcal{X}\}\right\}. \qquad (4.10)
\end{aligned}$$

The first term on the right hand side above is the loss due to price experimentation, (iv), whereas the second term is the sum of losses due to the other three sources (i-iii). Now, let us decompose the second term to see the tradeoff between (i), (ii) and (iii):

$$\begin{aligned}
\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{\sum_{t=1}^{T}\left(1-\frac{r(p_t,\theta_t)}{r^*(\theta_t)}\right)\mathbb{I}\{t\notin\mathcal{X}\}\right\} &= \mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{\sum_{j=0}^{\mathcal{C}}\sum_{s=n\tau^*_j+1}^{n\tau^*_{j+1}}\left(1-\frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}\{s\notin\mathcal{X}\}\right\}\\
&= \sum_{j=0}^{\mathcal{C}}\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{\sum_{s=n\tau^*_j+1}^{n\hat{\tau}^+_j}\left(1-\frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}\{s\notin\mathcal{X}\}\right.\\
&\qquad + \sum_{s=n\hat{\tau}^+_j+1}^{n\hat{\tau}^-_j}\left(1-\frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}\{s\notin\mathcal{X}\}\\
&\qquad \left.+ \sum_{s=n\hat{\tau}^-_j+1}^{n\tau^*_{j+1}}\left(1-\frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}\{s\notin\mathcal{X}\}\right\}. \qquad (4.11)
\end{aligned}$$

The first, second, and third sums inside the expectation on the right hand side above are the losses due to delay of true detections, noise in estimation, and early false alarms, respectively. Our next task is to find upper bounds on these sums. In the analysis of the losses associated with delayed correct detections (or early false alarms), the following lemma is key.

Lemma 3. (polynomial decay of detection error) *Let π be $D(\eta,\kappa,x_1,x_2)$ where $\kappa = c_\epsilon/(\eta\wedge\eta^2)$ and c_ϵ is a finite positive constant characterized by the distribution of $\{\epsilon_t\}$. For all i and k, let $\bar{\epsilon}_{ik} := m^{-1}\sum_{t\in\mathcal{X}_{ik}}\epsilon_t$, with $m=\lceil\kappa\log T\rceil$. Then,*

$$\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\left\{\left|\bar{\epsilon}_{ik}\right|\ge\tfrac{1}{2}\eta\right\}\le 2T^{-3/2}, \qquad (4.12)$$

for all $T\ge 3$, $\boldsymbol{\theta}\in\Theta^T$, $i=1,2$, and $k=0,1,2,\ldots,\lfloor T/n\rfloor$.

Remark 2. The constant c_ϵ, which appears in the above lemma as well as Lemmas 4, 5, and Theorem 4 below, is independent of T, B, δ, and completely characterized by the exponential moment condition, $\mathbb{E}[\exp(x\epsilon_t)] < \infty$ for all $|x| \le x_0$. For example, in the case $\epsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0,\sigma^2)$, we have $c_\epsilon = 12\sigma^2$. A general expression of c_ϵ is provided in the first paragraph in the proof of Lemma 3. Using Lemma 3, we first obtain the following upper bound on the loss due to detection delay.

Lemma 4. (loss due to delay of true detections) *Let π be $D(\eta,\kappa,x_1,x_2)$ with $\eta = \frac{|x_1-x_2|}{8(1\vee x_1 \vee x_2)}\delta$ and $\kappa = c_\epsilon/(\eta\wedge\eta^2)$, where c_ϵ is the constant given in Lemma 3, and $\vee$ denotes the maximum of two numbers. Then there exists a finite positive constant C_1 such that*

$$\mathbb{E}^\pi_{\boldsymbol{\theta}}\left\{\sum_{s=n\tau^*_j+1}^{n\hat{\tau}^+_j}\left(1-\frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}\{s\notin\mathcal{X}\}\right\} \;\le\; C_1\sqrt{T} \tag{4.13}$$

for all $T\ge 3$ and $\boldsymbol{\theta}\in\mathcal{V}(T,B,\delta)$.

In contrast to results in the single change-point detection literature, where the only uncertainty is about the time of change, the preceding lemma is proven without prior knowledge of the environment pre- and post-change. In this lemma, to estimate the expected loss due to detection delay we first analyze the detection test in (4.6), which repeatedly compares the average demand estimates in the current cycle k with the ones in cycles $L(k),\dots,k-1$, where $L(k)=\max\{\tau\le k:\chi_\tau=1\}$ denotes the latest detection cycle before k. As long as the demand parameter vector in cycle τ^*_j+1 is significantly different than one of the demand parameter vectors in cycles $L(\tau^*_j),\dots,\tau^*_j$, there is a high probability of detecting the j^{th} change-point. But, if almost all of the demand parameter vectors in cycles $L(\tau^*_j),\dots,\tau^*_j$ are the same as the demand parameter vector in cycle τ^*_j+1, this means that the unknown demand parameter sequence must have switched back to a value that was prevalent in the cycles that occurred after $L(\tau^*_j)$. In that case, it is unlikely that the detection test (4.6) will identify the j^{th} change-point, but this would not lead to substantial loss because almost all of the information accumulated since cycle $L(\tau^*_j)$ will be relevant in cycle τ^*_j+1. We formalize this argument in the proof of Lemma 4, and show that the loss due to detection delay is of order $\sqrt{T}$ under our detection policy.

Our next result builds on Lemma 3 to show that the loss due to false alarms is of order $\sqrt{T}$.

Lemma 5. (loss due to false alarms) *Let π be $D(\eta,\kappa,x_1,x_2)$ with $\eta = \frac{|x_1-x_2|}{8(1\vee x_1\vee x_2)}\delta$ and $\kappa = c_\epsilon/(\eta\wedge\eta^2)$, where c_ϵ is the constant given in Lemma 3. Then there exists a finite positive constant C_2 such that*

$$\mathbb{E}^\pi_{\boldsymbol{\theta}}\left\{\sum_{s=n\hat{\tau}^-_j+1}^{n\tau^*_{j+1}}\left(1-\frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}\{s\notin\mathcal{X}\}\right\} \;\le\; C_2\sqrt{T} \tag{4.14}$$

for all $T\ge 3$ and $\boldsymbol{\theta}\in\mathcal{V}(T,B,\delta)$.

It is worth noting that the setting studied in this section might include multiple false alarms because there is potentially more than one bursty change, and accordingly, the detection test (4.6) is repeated throughout the time horizon. Lemma 5 provides an upper bound on the revenue loss because of all such false alarms between the j^{th} and $(j+1)^{\text{st}}$ change-points.

Having found $O(\sqrt{T})$ upper bounds on the losses due to false detections, we prove in the following lemma; the loss due to estimation noise is also $O(\sqrt{T})$.

Lemma 6. (loss due to estimation noise) *Let π be $D(\eta,\kappa,x_1,x_2)$. Then there exists a finite positive constant C_3 such that*

$$\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{\sum_{s=n\hat{\tau}_j^{+}+1}^{n\hat{\tau}_j^{-}}\left(1-\frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}\{s\notin\mathcal{X}\}\right\} \;\le\; C_3\sqrt{T} \tag{4.15}$$

for all $T\geq 3$ and $\boldsymbol{\theta}\in\Theta^T$.

The preceding lemma provides an upper bound on the revenue loss incurred between the true detection after the j^{th} change-point and the first false detection before the $(j+1)^{\text{st}}$ change-point. During this time interval, there are no changes in demand parameters and no detections, meaning that there is no estimation inaccuracy due to changes, and the revenue loss is entirely caused by estimation error due to noisy demand observations. By a straightforward modification of Lemma 2, the price experimentation scheme in (4.3-4.4) implies that the loss due to estimation noise in this case is at most of order $\sqrt{T}$.

In the final result of this section, we combine Lemmas 4, 5, and 6 with inequality (4.11) to obtain the following performance bound.

Theorem 4. (near-optimality of the pricing-detection policy) *Let π be $D(\eta,\kappa,x_1,x_2)$ with $\eta=\frac{|x_1-x_2|}{8(1\vee x_1\vee x_2)}\delta$ and $\kappa=c_\epsilon/(\eta\wedge\eta^2)$, where c_ϵ is the constant given in Lemma 3. Then there exists a finite positive constant C such that $\mathcal{R}^{\pi}(T,B,\delta)\leq C\,T^{1/2}\log T$ for all $T\geq 3$.*

Remark 3. We note that our detection policy, $D(\eta,\kappa,x_1,x_2)$, uses the knowledge of δ in the choice of parameters η and κ, but does not require the knowledge of B. The constant C in the preceding theorem is $O(\delta^{-2})$ as $\delta\downarrow 0$. Note that as $\delta\downarrow 0$ the case of bursty changes converges to the case of smooth changes, and in the limit, the performance guarantee in Theorem 2 will eventually become tighter than the one in Theorem 4.

According to Theorem 4, the T-period regret of our detection policy is $O(T^{1/2}\log T)$. To put this result in perspective, we refer readers to two existing lower bounds: Keskin and Zeevi [22] derive a lower bound of order $T^{1/2}$ in a learning-and-earning problem in a static demand environment. Besbes and Zeevi [8] obtain another lower bound of order $T^{1/2}$ in a single change-point detection problem in which demand curves before and after the change-point are known. In light of these results, the policy in Theorem 4 is near-optimal in order (up to logarithmic terms in T).

In addition to the aforementioned lower bounds, Keskin and Zeevi [22] and Besbes and Zeevi [8] also provide policies that have $O(T^{1/2}\log T)$ regret in their settings, but because they focus on either learning or detection in isolation, they do not address the challenges arising in simultaneous learning and detection, such as the occurrence of a change-point prior to forming an informative (i.e., not-too-noisy) estimate of the average demand reference point for the detection test. To address such challenges, we employ a repeated detection test with a carefully chosen frequency to obtain the key results in Lemmas 3 and 4. Interestingly, the policy we design achieves essentially the same regret performance as in Keskin and Zeevi [22] and Besbes and Zeevi [8], this despite the fact that in our setting the problem facing the seller is more challenging compared to the formulations studied in the aforementioned papers.

5. Rapidly changing demand environments.

We now generalize the constant-budget problem formulation in Section 2 to more rapidly changing environments, where the change budget given in condition (2.5) is increasing in T. To be precise, take $\nu\in[0,1]$, and assume that θ_t are chosen from the rectangle $\Theta\subseteq\mathbb{R}\times\mathbb{R}_-$ such that

$$V_{\boldsymbol{\theta}}(T)\;\leq\;B\,T^{\nu}\qquad\text{for } T=1,2,\ldots \tag{5.1}$$

where $B>0$, and $V_{\boldsymbol{\theta}}(T)$ is the quadratic variation of $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots)$ in T periods, defined in (2.4). In condition (5.1), the parameter ν represents the *volatility* of the changing demand environment: if $\nu=0$ then we have the constant-budget problem formulation studied in preceding sections, whereas if $\nu=1$ then the demand environment is extremely volatile in the sense that there can be a substantial change in every single period. For intermediate values of $\nu\in(0,1)$, we obtain a spectrum of demand environments where the scale of change is characterized by the volatility parameter ν.

We incorporate the change budget in (5.1) into our performance metric as follows: let

$$\mathcal{R}^{\pi}(T,\nu) \;=\; \sup\big\{\Delta^{\pi}_{\boldsymbol{\theta}}(T):\boldsymbol{\theta}\in\mathcal{V}(T,\nu)\big\}, \tag{5.2}$$

where $\Delta^{\pi}_{\boldsymbol{\theta}}(T)$ is as defined in (2.9), and $\mathcal{V}(T,\nu)=\{\boldsymbol{\theta}:V_{\boldsymbol{\theta}}(T)\le BT^{\nu}\}$. Here we note that $\mathcal{V}(T,\nu)$ is a superset of $\mathcal{V}(T,B)$, namely the set of admissible demand parameter sequences in the constant-budget problem. The main question we address in this section is how much the regret would increase when we expand the set of admissible demand parameter sequences. We have the following lower bound on regret under condition (5.1).

Theorem 5. (lower bound on regret) *There exists a finite positive constant c such that $\mathcal{R}^{\pi}(T,\nu)\ge cT^{(2+\nu)/3}$ for any pricing policy π and time horizon $T\ge 3$.*

Note that when $\nu=1$ the revenue losses must grow linearly with the horizon, namely, the regret is no longer sublinear, and *no policy* is long-run-average optimal. To achieve the growth rate of regret in Theorem 5, we modify the moving window and decaying weights policies in Section 3.3 as follows.

First-order optimal policies in rapidly changing environments. As before, we consider policies that conduct price tests with a certain frequency, but due to increased volatility of the demand vector sequence $\boldsymbol{\theta}$, the frequency of price tests needs to be higher. That is, we let $n:=\lceil\kappa T^{(1-\nu)/3}\rceil\ge 2$ where κ is a scale parameter, and construct the set of periods at which the test prices will be charged as $\mathcal{X}_i:=\big\{t=kn+i\;:\;k=0,1,2,\ldots,\lfloor T/n\rfloor\big\}$ for $i=1,2$. Choosing two distinct test prices x_1 and x_2 in $[\ell,u]$, we let the price in period t be

$$p_t=\begin{cases} x_1 & \text{if } t\in\mathcal{X}_1\\ x_2 & \text{if } t\in\mathcal{X}_2\\ \varphi(\vartheta_t) & \text{otherwise,}\end{cases} \tag{5.3}$$

where ϑ_t is the truncated estimate of θ_t, which satisfies $\vartheta_t:=\arg\min_{\vartheta\in\Theta}\{\|\vartheta-\hat{\theta}_t\|\}$. The frequency of price tests in the experimentation scheme (5.3) is $2/n$, namely of order $T^{-(1-\nu)/3}$.

In rapidly changing environments, the moving window policy needs to have a smaller window size, whereas the decaying weights policy needs to have a more significant rate of decay. In the constant-budget problem, the window size of $M(\kappa,x_1,x_2)$ was of order $T^{2/3}$. In this section, we choose a window size of order $T^{(2-2\nu)/3}$: under condition (5.1), the *rapidly moving window policy* with parameters κ,x_1,x_2, denoted by $M_\nu(\kappa,x_1,x_2)$, chooses prices according to (5.3), and uses the weights $w^t_s=\mathbb{I}\{s\in\mathcal{X},\,s\ge t-n^2\}$ for $1\le s\le t$, where $\mathcal{X}=\mathcal{X}_1\cup\mathcal{X}_2$ and $n=\lceil\kappa T^{(1-\nu)/3}\rceil$. Similarly, the *rapidly decaying weights policy* with parameters μ,κ,x_1,x_2, denoted by $W_\nu(\mu,\kappa,x_1,x_2)$, chooses prices according to (5.3), and uses the weights $w^t_s=\big(1-\frac{t-\tilde{s}}{n^2}+\frac{(t-\tilde{s})^{1-\mu}}{n^2_\mu}\big)^{1/\mu}_+\mathbb{I}\{s\in\mathcal{X}\}$ for $1\le s\le t$, where $0<\mu\le 1$, $n=\lceil\kappa T^{(1-\nu)/3}\rceil$, and $n_\mu=nT^{\mu\nu}$.

In our next result, we extend Lemma 1 to the case of rapidly changing demand environments.

Lemma 7. (upper bound on aggregate estimation inaccuracy) *There exists a finite positive constant c_1, such that under either $M_\nu(\kappa,x_1,x_2)$ or $W_\nu(\mu,\kappa,x_1,x_2)$*

$$\sum_{t=n^2}^{T-1}\big\|\big(\mathcal{J}^t_t\big)^{-1}\mathcal{W}^t_t\big\|^2\le c_1T^{(2+\nu)/3} \tag{5.4}$$

for all $T = 1, 2, \ldots$ *and* $\boldsymbol{\theta} \in \mathcal{V}(T, \nu)$.

By Lemma 7 and a straightforward modification of Lemma 2, we generalize Theorem 2, and derive the following upper bound on the regret for $M_\nu(\kappa, x_1, x_2)$ and $W_\nu(\mu, \kappa, x_1, x_2)$.

Theorem 6. (first-order optimality) *Let* π *be either* $M_\nu(\kappa, x_1, x_2)$ *or* $W_\nu(\mu, \kappa, x_1, x_2)$. *Then there exist a finite positive constant* C *such that* $\mathcal{R}^\pi(T, \nu) \le C\, T^{(2+\nu)/3}$ *for all* $T \ge 3$.

Remark 4. The constant C in the preceding theorem is linear in B and independent of ν.

The preceding theorem provides a range of results for different degrees of change scales, quantifying how the volatility parameter ν influences the growth rate of regret. As ν increases, nature can cause larger estimation inaccuracy, and in response the seller needs to depreciate information faster, either by choosing a smaller moving window size, or by faster weight decay. Roughly speaking, every quanta of $O(T^\nu)$ in the change budget translates to an $O(T^{\nu/3})$ of regret.

6. Concluding remarks.

6.1. Discussion of main findings.

Measuring information depreciation. To compute the near-optimal information-depreciation rates in the settings analyzed in this paper, let us compare the sizes of the moving windows we constructed in these settings. In a static environment, we can use all past data within the entire time horizon; hence the size of the "moving" window is $O(T)$. In changing environments, we use moving windows of smaller order, such as the $O(T^{2/3})$ moving windows in Section 3. Given a particular demand environment, let the *information depreciation factor* be the ratio of the near-optimal moving window size in that environment to the nominal time horizon T. In static settings, the information depreciation factor is of order 1 by definition, and in the time-varying settings of Sections 3 and 5, the information depreciation factors of our policies are of order $T^{-1/3}$ and $T^{-(1+2\nu)/3}$, respectively. In the case of bursty changes, the information is depreciated only when a change-point is detected. Because the number of change-points is bounded by a constant independent of T in this case, our detection policy would rarely depreciate its information, and its information depreciation factor would be of order 1 (except when there are two closely timed change-points).

Structure of well-performing policies. Our study presents three families of dynamic pricing policies designed to perform well in changing demand environments. The moving window and decaying weights policies in Section 3 are based on a weighted least squares estimator that discounts older observations at a certain rate. The detection policy in Section 4 uses the same weighted least squares estimator, but can reduce the weight of all past observations to zero upon detecting a change. All of these policies have near-optimal performance in their respective settings, but at the same time they use quite distinct rules for weighing past observations, which suggests that successful pricing policies in presence of smooth and bursty changes can have very different structures.

Calibrating the volatility parameter. To design successful dynamic pricing policies in rapidly changing environments, a seller needs to characterize the volatility in the demand environment, which is represented by parameter ν in the problem formulation in Section 5. The demand volatility can be characterized by first observing the demand response to a given incumbent price $\hat{p}$ over N periods, and then measuring the average variation in expected demand as $v_N = \frac{1}{N}\sum_{t=1}^{N}(D_t - D_{t-1})^2 - 2\sigma^2$. Note that v_N estimates the average quadratic variation over N periods. In a demand environment with volatility parameter ν, v_N would be of order $N^\nu/N = N^{\nu-1}$. In light of this knowledge, the seller can run an ordinary least squares regression between $\log v_N$ and $\log N$, and calibrate the volatility parameter ν. To demonstrate how this could be done in practice, consider an example where the seller has access to demand observations under the prices 1.0, 1.1, …, 1.7.

Table 1. Numerical example for calibrating the volatility parameter*

$\hat{p}$	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7
N	41	96	38	27	35	43	44	66
v_N	0.148	0.123	0.219	0.306	0.347	0.327	0.371	0.356

* In each column, the seller computes v_N based on the demand sample $\{D_t, t=1,2,\ldots,N\}$ observed under price $\hat{p}$. The demand samples used for constructing this table are randomly generated in an environment that changes in a piecewise-linear cyclical pattern, which will be explicitly expressed in Section 6.3.

Given the demand samples observed under each price, suppose that the sample sizes N and the observed values of the statistic v_N are as in Table 1.

In this example, running a regression for the relationship $\log(v_N) \approx \zeta_0 + \zeta_1 \log N$ would result in the estimates $(\hat{\zeta}_0, \hat{\zeta}_1) = (0.525, -0.495)$. Thus, the calibrated value of the volatility parameter is $\hat{\nu} = 1 + \hat{\zeta}_1 = 0.505$. The true value of the volatility parameter is $\nu = 0.5$ in this example.

Linear demand assumption and asymptotically optimal semi-myopic policies. Linear regression models are commonly used in econometrics to express reduced-form relationships between variables. In this paper, we model the relationship between price and demand in a similar linear fashion, but it is possible to extend our analysis to the case of a generalized linear model of the form $d = f(L(p))$, where d is the expected demand under price p, L is a linear function with unknown parameters, and f is a general "link function" known with certainty. In the case of a linear demand model, ordinary least squares regression is usually used as a special case of maximum likelihood estimation, whereas in the extension to generalized linear models one can use maximum quasi-likelihood estimation in a similar way (see den Boer and Zwart [14, 15]). This generalization involves substantially more technical detail, while leading to essentially the same conclusions in terms of estimation performance as shown by den Boer and Zwart [14, 15]. To achieve a more transparent analysis and to obtain a wider variety of results, we have chosen to use a linear demand model. In practice the relationship between price and demand can be described by more general functional forms, in which expected demand is a smooth decreasing function of price. In such cases, the linear demand assumption leads to model misspecification, but as shown by Besbes and Zeevi [9], such model misspecification can be mitigated by designing asymptotically optimal semi-myopic policies under the linear demand assumption. In essence, their argument is that as long as the expected demand function has a well-defined derivative around the optimal price, one can use a linear approximation to the expected demand function within a small neighborhood of the optimal price without incurring substantial loss. The essential principle behind asymptotically optimal semi-myopic policies, which are also employed in this paper as well as in Keskin and Zeevi [22] and Besbes and Zeevi [9], is using price experimentation to ensure that the myopic price is within a small neighborhood of the optimal price with high probability, and thereby limiting losses due to model misspecification.

Data storage constraints. Moving window and decaying weights policies designed in Section 3 differ markedly in how they store price and sales data. While the decaying weights policy makes use of all historical observations, the moving window requires only a relatively small number of observations. Hence, data storage considerations favor moving window policies.

6.2. Relation to other stochastic optimization approaches.

Our problem formulation is related to various sequential stochastic optimization formulations, but unlike our work the vast majority of that literature has focused on static environments where the underlying objective function does not change over time. We note that this literature contains parameterized bandit formulations that can potentially be applied to pricing contexts [see, e.g., 27], but to the best of our knowledge these formulations also restrict attention to static environments.

Concurrently with our work, Besbes et al. [6] has focused on the exploration-exploitation tradeoff in a stochastic approximation problem with gradually changing environments. Our study significantly differs from theirs in the following ways: first, unlike Besbes et al. [6] we study the case where the variation budget B is unknown to the seller, which allows us to quantify the additional cost of adapting to an unknown variation budget (see Sections 3.3 and 3.4). Second, our work analyzes both gradually and abruptly changing environments. As explained earlier, this helps us identify a stark contrast between these two environments (see our discussion in Section 1.2). Third, we study how to design bona fide tracking policies based on (1) moving windows, (2) decaying weights, and (3) joint detection-and-estimation in various changing environments, and provide a mathematical analysis to characterize the performance of our policies, which is not present in antecedent work. Fourth, due to the aforementioned differences between studied policies, the $O(T^{2/3})$ and $O(T^{1/2}\log T)$ regret bounds in our paper are obtained through entirely independent techniques than the regret bounds in Besbes et al. [6]. To be precise, we derive our regret bounds via a combination of eigenanalysis and concentration inequalities for a family of vector martingales as opposed to Besbes et al. [6] who essentially rely on known results for online gradient descent in an adversarial setting which are modified to work well in the changing environment they consider. Finally, there is a fundamental difference between the main messages of our work and that of Besbes et al. [6]. Besbes et al. [6] focus on repetitively restarting a policy with a predetermined frequency to achieve a performance guarantee in gradually changing environments. In contrast, we use restarting only in an adaptive manner if the changes are abrupt. If the changes are smooth, then we employ smooth information-depreciation policies (such as moving windows or decaying weights). In particular, we do not prescribe using restarting if the demand environment is smoothly changing.

6.3. Numerical example. To demonstrate the practical implementability of our policies, we simulate their performance in a numerical example. Suppose the feasible set of demand parameters is $\Theta = [100,120]\times[-50,-35]$, and the demand parameter sequence follows a piecewise-linear cyclical pattern as follows: let $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_T)$ be such that $\theta_1 = (110,-49.25)$ and

$$\theta_{t+1} - \theta_t \;=\; \begin{cases} +(0,T^{-1/2}) \text{ if } \; t \le K \pmod{2K} \\ -(0,T^{-1/2}) \text{ otherwise,} \end{cases} \tag{6.1}$$

for $t = 1,2,\ldots,T-1$, where $K = \lceil T^{2/3}\rceil$. Note that, given any time horizon T, the aforementioned parameter sequence $\boldsymbol{\theta}$ satisfies $V_{\boldsymbol{\theta}}(T) \le 1$; that is, $B = 1$ for the class of parameter sequences described above. Assume that standard deviation of demand shocks is $\sigma = 1$, and the feasible price range is $[\ell,u] = [0.9,1.8]$.

We consider four different policies in this setting. The first two are the moving window and decaying weights policies defined in Section 3.3. The third one is the fixed-step stochastic approximation policy, hereafter abbreviated as *fixed-step SA*, which is usually considered as a heuristic policy for stochastic optimization problems in changing environments [see 5, Chapter 4]. The fourth policy is the restarting stochastic approximation policy, abbreviated as *restarting SA*, which was proposed by Besbes et al. [6, Section 5.2, EGS algorithm]. The fixed-step SA and restarting SA belong to a broad class of stochastic approximation policies that utilize the noisy observations on the revenue curve to first estimate the gradient of this curve, and then move the price in the gradient direction to find the revenue-maximizing price [a detailed description of this gradient estimation can be found in Section 3.3 of 9]. Stochastic approximation policies designed for static environments typically decay the step size for the moves in the gradient direction [see, e.g., 24]. In contrast, fixed-step SA uses a constant step-size sequence so that the moves in the gradient direction do not diminish in a changing environment. Restarting SA keeps "rebooting" the stochastic approximation routine with a pre-determined frequency in an open-loop fashion, and the step-size sequence is reset at each such epoch.

As displayed in Figure 1, the moving window and decaying weights policies significantly outperform fixed-step SA and restarting SA in the numerical example described above.

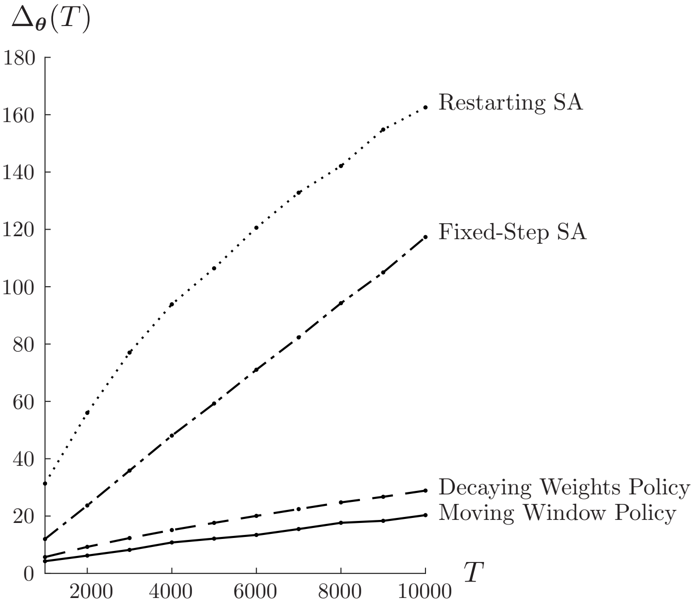

FIGURE 1. REGRET COMPARISON. The four curves display the T-period regret of the moving window policy (solid curve), the decaying weights policy (dashed curve), fixed-step SA (dash-dotted curve), and restarting SA (dotted curve). For the moving window and decaying weights policies, the scale parameter is $\kappa = 0.5$ and the experimental prices are $x_1 = 1.1$ and $x_2 = 1.3$. The decay parameter of decaying weights policy is $\mu = 0.5$. The initial price of fixed-step SA and restarting SA is $x_1 = 1.3$. The step size parameters of restarting SA (described on p. 11 of Besbes et al. [6]) are $d = 0.1$ and $H = 1$. The corresponding step size parameters for fixed-step SA are $d = 0.0002$ and $H = 1$.

We note that the estimated growth rate of regret is close to $T^{2/3}$ for the moving window, decaying weights, and restarting SA policies, but for fixed-step SA, we estimate the growth rate of regret to be approximately T. To be more precise, a log-log regression reveals that the T-period regret of the moving window, decaying weights, fixed-step SA, and restarting SA policies are in the order of $T^{0.68}$, $T^{0.69}$, $T^{0.98}$, and $T^{0.73}$, respectively (with $R^2 = 0.98$; all regression coefficients are statistically significant with $p < 0.001$).

The superior performance of the moving window and decaying weights policies is due to the fact that they always maintain memory of the recently obtained information. Restarting SA forgets all historical information in an open-loop manner, which makes this policy repeatedly incur losses due to initializations. Fixed-step SA displays better performance than restarting SA over short time horizons, but given its (asymptotically) linearly growing regret, the performance of fixed-step SA is expected to worsen over longer time horizons.

Appendix A: Proof of Theorem 1. Divide the time horizon into cycles of $N = \lceil k_0 T^{2/3} \rceil$ periods, where $k_0 = 4^{2/3} B^{-2/3}$, and consider the setting in which (i) $\epsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, (ii) the value of θ_t can change only in the first period of a cycle, and (iii) θ_t takes values in the set $\{y_0, y_1\}$, where $y_0 = (a_0, b_0) = (2, -1)$ and $y_1 = (a_1, b_1) = (2 + N^{-1/4}, -1 - N^{-1/4})$. Note that (ii), (iii), and the above choice of N imply that

$$V_{\boldsymbol{\theta}}(T) \;\le\; \left(\frac{T}{N} + 1\right) \|y_0 - y_1\|^2 \;\overset{\text{(a)}}{\le}\; \frac{4T}{N^{3/2}} \;\le\; B, \tag{A.1}$$

where: (a) follows because $N \le T$ and $\|y_0 - y_1\|^2 = 2N^{-1/2}$. Therefore the setting described above satisfies the quadratic variation bound in (2.5).

Now, focus on a single cycle, which is composed of N periods. Let $\mathbf{P}_i^\pi$ be a probability measure satisfying

$$\mathbf{P}_i^\pi(D_1 \le \xi_1, \ldots, D_N \le \xi_N) \;=\; \prod_{t=1}^{N} \Phi\left(\frac{\xi_t - a_i - b_i p_t}{\sigma}\right) \qquad \text{for } \xi_1, \ldots, \xi_T \in \mathbb{R}, \tag{A.2}$$

where $\Phi(\cdot)$ denotes the standard Gaussian cumulative distribution function, and $p=(p_1,p_2,\ldots)$ is the price sequence formed under policy π and demand realization $D=(D_1,D_2,\ldots)$. Then, the Kullback-Leibler divergence from $\mathbf{P}_0^\pi$ to $\mathbf{P}_1^\pi$ is

$$\mathcal{K}(\mathbf{P}_0^\pi,\mathbf{P}_1^\pi) \;:=\; \mathbf{E}_0^\pi \log\left(\frac{\displaystyle\prod_{t=1}^{N}\phi\left(\frac{D_t-a_0-b_0p_t}{\sigma}\right)}{\displaystyle\prod_{t=1}^{N}\phi\left(\frac{D_t-a_1-b_1p_t}{\sigma}\right)}\right), \tag{A.3}$$

where $\mathbf{E}_0^\pi$ is the expectation operator associated with $\mathbf{P}_0^\pi$, and $\phi(\cdot)$ denotes the standard Gaussian density. By elementary algebra, we can re-express (A.3) as follows:

$$\begin{aligned}\mathcal{K}(\mathbf{P}_0^\pi,\mathbf{P}_1^\pi) &= -\frac{1}{2\sigma^2}\,\mathbf{E}_0^\pi\left\{\sum_{t=1}^{N}\left[(D_t-a_0-b_0p_t)^2-(D_t-a_1-b_1p_t)^2\right]\right\}\\ &= -\frac{1}{2\sigma^2}\,\mathbf{E}_0^\pi\left\{\sum_{t=1}^{N}\left[\epsilon_t^2-\big(\epsilon_t+a_0-a_1+(b_0-b_1)p_t\big)^2\right]\right\},\end{aligned} \tag{A.4}$$

because $D_t=a_0+b_0p_t+\epsilon_t$ under $\mathbf{P}_0^\pi$. Let $\delta=y_0-y_1$ and $X_t=\left[\begin{smallmatrix}1\\ p_t\end{smallmatrix}\right]$. Then, the preceding identity becomes

$$\begin{aligned}\mathcal{K}(\mathbf{P}_0^\pi,\mathbf{P}_1^\pi) &= -\frac{1}{2\sigma^2}\,\mathbf{E}_0^\pi\left\{\sum_{t=1}^{N}\left[\epsilon_t^2-(\epsilon_t-\delta\cdot X_t)^2\right]\right\}\\ &= -\frac{1}{2\sigma^2}\,\mathbf{E}_0^\pi\left\{\sum_{t=1}^{N}(2\epsilon_t-\delta\cdot X_t)\;\delta\cdot X_t\right\}\\ &\overset{(b)}{=} \frac{1}{2\sigma^2}\,\mathbf{E}_0^\pi\left\{\sum_{t=1}^{N}(\delta\cdot X_t)^2\right\}\\ &\overset{(c)}{=} \frac{1}{2\sigma^2}\,\mathbf{E}_0^\pi\left\{\sum_{t=1}^{N}N^{-1/2}\big(p_t-1\big)^2\right\}\\ &\overset{(d)}{=} \frac{1}{2\sigma^2N^{1/2}}\,\mathbf{E}_0^\pi\left\{\sum_{t=1}^{N}\big(p_t-\varphi(y_0)\big)^2\right\}\\ &\overset{(e)}{=} \frac{1}{2\sigma^2N^{1/2}}\,\boldsymbol{\Delta}_0^\pi(N),\end{aligned} \tag{A.5}$$

where: $\boldsymbol{\Delta}_i^\pi(N)$ denotes the N-period regret given that policy π is exercised and $\theta_t=y_i$ for all $t=1,\ldots,N$ and $i=0,1$, (b) follows because the ϵ_t are independent and have zero mean, (c) follows because $\delta=(-N^{-1/4},N^{1/4})$, (d) follows because $\varphi(y_0)=1$, and (e) follows by the definition of regret in (2.9) and the fact that $b_0/r^*(y_0)=-1$.

We will consider two cases for the value of $\mathcal{K}(\mathbf{P}_0^\pi,\mathbf{P}_1^\pi)$. Let $\eta>0$.

Case 1. $\mathcal{K}(\mathbf{P}_0^\pi,\mathbf{P}_1^\pi)>\eta$. By (A.5), we deduce that

$$\boldsymbol{\Delta}_0^\pi(N)\geq 2\sigma^2\eta N^{1/2}. \tag{A.6}$$

Case 2. $\mathcal{K}(\mathbf{P}_0^\pi,\mathbf{P}_1^\pi)\leq\eta$. Define $I_i:=\left[\varphi(y_i)-\tfrac{1}{4}N^{-1/4},\varphi(y_i)+\tfrac{1}{4}N^{-1/4}\right]$ for $i=0,1$, and let χ_t be a random variable such that

$$\chi_t=\begin{cases}1 \text{ if } p_t\in I_0\\ 0 \text{ otherwise,}\end{cases} \tag{A.7}$$

for all t. Then, we have

$$\begin{aligned}\boldsymbol{\Delta}_0^{\pi}(N)+\boldsymbol{\Delta}_1^{\pi}(N) &\geq \left(\frac{2b_0}{a_0}\right)^2\sum_{t=1}^{N}\mathbf{E}_0^{\pi}\big(p_t-\varphi(y_0)\big)^2+\left(\frac{2b_1}{a_1}\right)^2\sum_{t=1}^{N}\mathbf{E}_1^{\pi}\big(p_t-\varphi(y_1)\big)^2\\ &\overset{\text{(e)}}{\geq} k_1N^{-1/2}\sum_{t=1}^{N}\Big(\mathbf{P}_0^{\pi}(p_t\notin I_0)+\mathbf{P}_1^{\pi}(p_t\notin I_1)\Big)\\ &\overset{\text{(f)}}{\geq} k_1N^{-1/2}\sum_{t=1}^{N}\Big(\mathbf{P}_0^{\pi}(\chi_t=0)+\mathbf{P}_1^{\pi}(\chi_t=1)\Big),\end{aligned}\tag{A.8}$$

where: $k_1=\frac{1}{4}\min\big\{(b_0/a_0)^2,(b_1/a_1)^2\big\}$, (e) follows because $\big(p_t-\varphi(y_i)\big)^2>\frac{1}{16}N^{-1/2}$ a.s. on the event $\{p_t\notin I_i\}$ for $i=0,1$, and (f) follows because $p_t\notin I_1$ is implied by $\chi_t=1$. By Tsybakov's bound on minimax probability of error for two hypotheses [30, p. 90, Theorem 2.2(iii)], we know that $\mathcal{K}(\mathbf{P}_0^{\pi},\mathbf{P}_1^{\pi})\leq\eta$ implies $\mathbf{P}_0^{\pi}(\chi_t=0)+\mathbf{P}_1^{\pi}(\chi_t=1)\geq\frac{1}{4}\exp(-\eta)$. Therefore we deduce by (A.8) that

$$\max_{i=0,1}\{\boldsymbol{\Delta}_i^{\pi}(N)\}\geq\frac{1}{4}k_1\exp(-\eta)N^{1/2}.\tag{A.9}$$

Combining (A.6) and (A.9), we get $\max_{i=0,1}\{\boldsymbol{\Delta}_i^{\pi}(N)\}\geq k_2N^{1/2}$, where $k_2=\max\{2\sigma^2\eta,\frac{1}{4}k_1\exp(-\eta)\}$. Therefore we conclude that

$$\begin{aligned}\sup\big\{\Delta_{\boldsymbol{\theta}}^{\pi}(T):V_{\boldsymbol{\theta}}(T)\leq B\big\} &\overset{\text{(g)}}{\geq} \left\lfloor\frac{T}{N}\right\rfloor\max_{i=0,1}\{\boldsymbol{\Delta}_i^{\pi}(N)\}\\ &\geq k_2\left\lfloor\frac{T}{N}\right\rfloor N^{1/2}\\ &\geq \frac{1}{2}k_2N^{-1/2}T\\ &\geq c\,B^{1/3}T^{2/3},\end{aligned}\tag{A.10}$$

where: $c=\frac{1}{8}k_2$ and (g) follows because there are at least $\lfloor T/N\rfloor$ cycles in T periods. Q.E.D.

Appendix B: Proof of the results in Section 3.

Proof of Lemma 1. We will first prove (3.12) for $M(\kappa,x_1,x_2)$. For the choice of weights in (3.8), the estimation inaccuracy in period t is

$$\big(\mathcal{J}_t^t\big)^{-1}\mathcal{W}_t^t = \big(\mathcal{J}_t^t\big)^{-1}\sum_{s=1}^{t}w_s^tX_sX_s^{\mathsf{T}}(\theta_s-\theta_{t+1}) = \big(\mathcal{J}_t^t\big)^{-1}\sum_{\substack{s\in\mathcal{X}_1\cup\mathcal{X}_2\\ t-n^2\leq s\leq t}}X_sX_s^{\mathsf{T}}(\theta_s-\theta_{t+1}),\tag{B.1}$$

for all t. Now note that $\big(\mathcal{J}_t^t\big)^{-1}=\big(\mathcal{I}_t^t\big)^{-1}\mathfrak{X}^{-1}=n^{-1}\mathfrak{X}^{-1}$ for all $t\geq n^2$ under $M(\kappa,x_1,x_2)$. Moreover, for $i=1,2$, we have $X_sX_s^{\mathsf{T}}=\left[\begin{smallmatrix}1 & x_i\\ x_i & x_i^2\end{smallmatrix}\right]$ for all $s\in\mathcal{X}_i$. Plugging these expressions of $\big(\mathcal{J}_t^t\big)^{-1}$ and $X_sX_s^{\mathsf{T}}$ into (B.1), we get

$$\begin{aligned}\big(\mathcal{J}_t^t\big)^{-1}\mathcal{W}_t^t = &\sum_{\substack{s\in\mathcal{X}_1\\ t-n^2\leq s\leq t}}\tfrac{1}{(x_1-x_2)n}\begin{bmatrix}-x_2 & -x_1x_2\\ 1 & x_1\end{bmatrix}(\theta_s-\theta_{t+1})\\ &+\sum_{\substack{s\in\mathcal{X}_2\\ t-n^2\leq s\leq t}}\tfrac{1}{(x_1-x_2)n}\begin{bmatrix}-x_1 & -x_1x_2\\ 1 & x_2\end{bmatrix}(\theta_s-\theta_{t+1}).\end{aligned}\tag{B.2}$$

Therefore we have

$$\begin{aligned}\left\|\left(\mathcal{J}_t^t\right)^{-1}\mathcal{W}_t^t\right\| &\overset{\text{(a)}}{\leq} \sum_{\substack{s\in\mathcal{X}_1\\ t-n^2\leq s\leq t}}\left\|\tfrac{1}{(x_1-x_2)n}\begin{bmatrix}-x_2 & -x_1x_2\\ 1 & x_1\end{bmatrix}(\theta_s-\theta_{t+1})\right\| + \sum_{\substack{s\in\mathcal{X}_2\\ t-n^2\leq s\leq t}}\left\|\tfrac{1}{(x_1-x_2)n}\begin{bmatrix}-x_1 & -x_1x_2\\ 1 & x_2\end{bmatrix}(\theta_s-\theta_{t+1})\right\|\\ &\overset{\text{(b)}}{\leq} \frac{1}{n}\sum_{\substack{s\in\mathcal{X}_1\cup\mathcal{X}_2\\ t-n^2\leq s\leq t}}\left\|\theta_s-\theta_{t+1}\right\|\\ &\leq 2\max_{t-n^2\leq s\leq t}\left\|\theta_s-\theta_{t+1}\right\|, \end{aligned} \tag{B.3}$$

for all $t\geq 2n$, where: (a) follows by (B.2) and triangle inequality, and (b) follows by the fact that the eigenvalues of $\frac{1}{(x_1-x_2)n}\begin{bmatrix}-x_2 & -x_1x_2\\ 1 & x_1\end{bmatrix}$ and $\frac{1}{(x_1-x_2)n}\begin{bmatrix}-x_1 & -x_1x_2\\ 1 & x_2\end{bmatrix}$ are 0 and $\pm n^{-1}$. Squaring and summing both sides of (B.3) over $t=n^2,\ldots,T-1$, we get

$$\begin{aligned}\sum_{t=n^2}^{T-1}\left\|\left(\mathcal{J}_t^t\right)^{-1}\mathcal{W}_t^t\right\|^2 &\leq 4\sum_{t=n^2}^{T-1}\max_{t-n^2\leq s\leq t}\left\|\theta_s-\theta_{t+1}\right\|^2\\ &\overset{\text{(c)}}{\leq} 4\sum_{j=1}^{\lceil T/n^2\rceil}\sum_{i=1}^{n^2}\max_{(j-1)n^2+i\leq s\leq jn^2+i}\left\|\theta_s-\theta_{jn^2+i+1}\right\|^2\\ &\overset{\text{(d)}}{=} 4\sum_{i=1}^{n^2}\sum_{j=1}^{\lceil T/n^2\rceil}\max_{(j-1)n^2+i\leq s\leq jn^2+i}\left\|\theta_s-\theta_{jn^2+i+1}\right\|^2\\ &\overset{\text{(e)}}{\leq} 4\sum_{i=1}^{n^2}V_{\boldsymbol{\theta}}(T) = 4n^2V_{\boldsymbol{\theta}}(T), \end{aligned} \tag{B.4}$$

where: (c) follows by expressing the time index as $t=jn^2+i$, (d) follows by changing the order of summation, and (e) follows by (2.4) because $\{t_j=jn^2+i: j=0,1,\ldots,\lceil T/n^2\rceil\}$ is a partition of $\{1,\ldots,T\}$ for all i. By (2.5) and the preceding inequality, we conclude that

$$\sum_{t=n^2}^{T-1}\left\|\left(\mathcal{J}_t^t\right)^{-1}\mathcal{W}_t^t\right\|^2 \leq 4n^2B \overset{\text{(f)}}{\leq} 16\kappa^2BT^{2/3}, \tag{B.5}$$

where (f) follows because $n=\lceil\kappa T^{1/3}\rceil$. We obtain (3.12) by letting $c_1=16\kappa^2B$.

Secondly, we prove (3.12) for $W(\mu,\kappa,x_1,x_2)$. Note that, because $\left(\mathcal{J}_t^t\right)^{-1}=\left(\mathcal{I}_t^t\right)^{-1}\mathfrak{X}^{-1}$ for all $t\geq n^2$, we have

$$\begin{aligned}\left(\mathcal{J}_t^t\right)^{-1}\mathcal{W}_t^t &= \sum_{s\in\mathcal{X}_1}\tfrac{w_s^t}{(x_1-x_2)\mathcal{I}_t^t}\begin{bmatrix}-x_2 & -x_1x_2\\ 1 & x_1\end{bmatrix}(\theta_s-\theta_{t+1})\\ &\quad+\sum_{s\in\mathcal{X}_2}\tfrac{w_s^t}{(x_1-x_2)\mathcal{I}_t^t}\begin{bmatrix}-x_1 & -x_1x_2\\ 1 & x_2\end{bmatrix}(\theta_s-\theta_{t+1}). \end{aligned} \tag{B.6}$$

Consequently, we get

$$\begin{aligned}\left\|\left(\mathcal{J}_t^t\right)^{-1}\mathcal{W}_t^t\right\| &\overset{\text{(a}'\text{)}}{\leq} \sum_{s\in\mathcal{X}_1}w_s^t\left\|\tfrac{1}{(x_1-x_2)n}\begin{bmatrix}-x_2 & -x_1x_2\\ 1 & x_1\end{bmatrix}(\theta_s-\theta_{t+1})\right\| + \sum_{s\in\mathcal{X}_2}w_s^t\left\|\tfrac{1}{(x_1-x_2)n}\begin{bmatrix}-x_1 & -x_1x_2\\ 1 & x_2\end{bmatrix}(\theta_s-\theta_{t+1})\right\|\\ &\overset{\text{(b}'\text{)}}{\leq} \left(\mathcal{I}_t^t\right)^{-1}\sum_{s\in\mathcal{X}_1\cup\mathcal{X}_2}w_s^t\left\|\theta_s-\theta_{t+1}\right\| \end{aligned}$$

$$
\begin{aligned}
&\overset{(\mathrm{c}')}{\leq} \big(\mathcal{I}_t^t\big)^{-1} n^{-2} \sum_{\substack{s\in\mathcal{X}_1\cup\mathcal{X}_2\\ 1\leq s<t-n^2}} \big\|\theta_s-\theta_{t+1}\big\| + \big(\mathcal{I}_t^t\big)^{-1} \sum_{\substack{s\in\mathcal{X}_1\cup\mathcal{X}_2\\ t-n^2\leq s\leq t}} \big\|\theta_s-\theta_{t+1}\big\| \\
&\leq \frac{2(t-n^2-1)}{n^3}\big(\mathcal{I}_t^t\big)^{-1} \max_{1\leq s<t-n^2} \big\|\theta_s-\theta_{t+1}\big\| \;+\; \frac{2n^2}{n}\big(\mathcal{I}_t^t\big)^{-1} \max_{t-n^2\leq s\leq t} \big\|\theta_s-\theta_{t+1}\big\| \\
&\leq \frac{2t}{n^3}\big(\mathcal{I}_t^t\big)^{-1} \max_{1\leq s<t-n^2} \big\|\theta_s-\theta_{t+1}\big\| \;+\; 2n\big(\mathcal{I}_t^t\big)^{-1} \max_{t-n^2\leq s\leq t} \big\|\theta_s-\theta_{t+1}\big\|, \qquad \text{(B.7)}
\end{aligned}
$$

for all $t\geq n^2$, where: (a′) follows by (B.6) and triangle inequality, (b′) follows by the fact that the eigenvalues of $\frac{1}{(x_1-x_2)\mathcal{I}_t^t}\begin{bmatrix} -x_2 & -x_1x_2 \\ 1 & x_1 \end{bmatrix}$ and $\frac{1}{(x_1-x_2)\mathcal{I}_t^t}\begin{bmatrix} -x_1 & -x_1x_2 \\ 1 & x_2 \end{bmatrix}$ are 0 and $\pm\big(\mathcal{I}_t^t\big)^{-1}$, and (c′) follows by (3.9) and the fact that $w_s^t\leq w_{t-n^2}^t\leq n^{-2}$ for all $s<t-n^2$, and $w_s^t\leq 1$ for all $s\leq t$. We square and sum both sides of (B.7) over $t=n^2,\ldots,T-1$ to obtain

$$
\begin{aligned}
\sum_{t=n^2}^{T-1} \big\|\big(\mathcal{J}_t^t\big)^{-1}\mathcal{W}_t^t\big\|^2 \;\leq\; & 8n^{-6}\sum_{t=n^2}^{T-1} t^2\big(\mathcal{I}_t^t\big)^{-2} \max_{1\leq s<t-n^2} \big\|\theta_s-\theta_{t+1}\big\|^2 \\
& + 8n^2 \sum_{t=n^2}^{T-1} \big(\mathcal{I}_t^t\big)^{-2} \max_{t-n^2\leq s\leq t} \big\|\theta_s-\theta_{t+1}\big\|^2. \qquad \text{(B.8)}
\end{aligned}
$$

Note that, for all $t\geq n^2$, we have

$$
\begin{aligned}
\mathcal{I}_t^t \;=\; \frac{1}{2}\sum_{s=1}^{t} w_s^t \;&\geq\; \frac{1}{2}\sum_{\substack{s\in\mathcal{X}\\ 1\leq s\leq t}} \left(1-\frac{t-s}{n^2}+\frac{(t-s)^{1-\mu}}{n^2}\right)_+^{\frac{1}{\mu}} \\
&\geq\; \frac{1}{2}\sum_{\substack{s\in\mathcal{X}\\ 1\leq s\leq t}} \left(1-\frac{1}{n}\left\lceil\frac{t-s}{n}\right\rceil\right)_+^{\frac{1}{\mu}} \\
&\overset{(\mathrm{d}')}{=}\; n^{-\frac{1}{\mu}}\sum_{k=1}^{n-1}(n-k)^{\frac{1}{\mu}} \\
&\geq\; n^{-\frac{1}{\mu}}\int_0^{n-1}\xi^{\frac{1}{\mu}}\,d\xi \\
&\geq\; c_\mu n\,, \qquad \text{(B.9)}
\end{aligned}
$$

where: $c_\mu=2^{-(\frac{1}{\mu}+1)}/\big(\frac{1}{\mu}+1\big)$, and (d′) follows by letting $k=\lceil(t-s)/n\rceil$. By the argument used to derive (B.4-B.5), and the fact that $\mathcal{I}_t^t\geq c_\mu n$ for all $t\geq n^2$, we deduce that the second term on the right hand side of (B.8) is bounded above by $32c_\mu^{-2}\kappa^2BT^{2/3}$. To find an upper bound on the first term, we note that

$$
\begin{aligned}
8n^{-6}\sum_{t=n^2+1}^{T-1} t^2\big(\mathcal{I}_t^t\big)^{-2} \max_{1\leq s<t-n^2}\big\|\theta_s-\theta_{t+1}\big\|^2 \;&\overset{(\mathrm{e}')}{\leq}\; 8n^{-6}\sum_{t=n^2+1}^{T-1} t^2\big(\mathcal{I}_t^t\big)^{-2}B \\
&\overset{(\mathrm{f}')}{\leq}\; 8c_\mu^{-2}n^{-8}\sum_{t=n^2+1}^{T-1} t^2B \\
&\leq\; 8c_\mu^{-2}n^{-8}BT^3, \qquad \text{(B.10)}
\end{aligned}
$$

where: (e′) follows by (2.5), and (f′) follows because $\mathcal{I}_t^t>c_\mu n$ for all $t\geq n^2$. Furthermore, because $n=\lceil\kappa T^{1/3}\rceil$, the right hand side of the preceding inequality is less than or equal to $8c_\mu^{-2}\kappa^{-8}BT^{1/3}$. Thus the right hand side of (B.8) is bounded above by $8c_\mu^{-2}(\kappa^{-8}+4\kappa^2)BT^{2/3}$.   Q.E.D.

Proof of Lemma 2. Because $\{\epsilon_t\}$ have a light-tailed distribution with mean zero and variance σ^2, we know by elementary real analysis that there exists a constant ν_0 such that $\mathbb{E}[\exp(x\epsilon_t)] \le \exp(\frac{1}{2}\nu_0\sigma^2 x^2)$ for all x satisfying $|x| \le x_0$ [see, e.g., 22, for a standard derivation of this constant]. For any given $t = 1, 2, \ldots$ and $y \in \mathbb{R}^2$ such that $\|y\| = z$, define a stochastic process $\{\mathcal{Z}_s^{y,t},\, s = 1, 2, \ldots\}$ such that $\mathcal{Z}_0^{y,t} = 1$ and

$$\mathcal{Z}_s^{y,t} = \begin{cases} \exp\left\{\frac{1}{\zeta}\left(y \cdot \mathcal{M}_s^t - \frac{1}{2} y^{\mathsf{T}} \mathcal{J}_s^t y\right)\right\} & \text{if } s \le t \\ \mathcal{Z}_{s-1}^{y,t} & \text{otherwise,} \end{cases} \tag{B.11}$$

where $\zeta = (1 \vee z)\big(\nu_0 \vee (x^*/x_0)\big)\sigma^2$ and $x^* = \max_{\|y\| \le 1,\, p \in [\ell,u]}\{|y_1 + y_2 p|/\sigma^2\}$. Let $\mathcal{F}_s := \sigma(\epsilon_1, \ldots, \epsilon_s)$. Using the tower property and the fact that $\mathcal{Z}_s^{y,t}$ is integrable for all s, we get

$$\mathbb{E}_{\boldsymbol{\theta}}^{\pi}[\mathcal{Z}_s^{y,t} | \mathcal{F}_{s-1}] \;=\; \exp\left\{\frac{1}{\zeta}\Big(y \cdot \mathcal{M}_{s-1}^t - \frac{1}{2} y^{\mathsf{T}} \mathcal{J}_s^t y\Big)\right\} \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big[\exp\Big\{\frac{1}{\zeta}\, y \cdot (\mathcal{M}_s^t - \mathcal{M}_{s-1}^t)\Big\}\Big| \mathcal{F}_{s-1}\Big],$$

for $s \le t$. To find an upper bound on the conditional expectation on the right hand side of the identity immediately above, note that $\mathcal{M}_s^t - \mathcal{M}_{s-1}^t = w_s^t X_s \epsilon_s$, and $|y \cdot (w_s^t X_s)|/\zeta = w_s^t\, |y_1 + y_2 p_s|/\zeta \le w_s^t\, |y_1 + y_2 p_s| x_0/(z x^* \sigma^2) \le w_s^t\, x_0 \le x_0$ for all $p_s \in [\ell, u]$, because $w_s^t \le 1$ for all $s \le t$ by definition of the weights in (3.16). As a result, the conditional expectation on the right hand side of the preceding identity satisfies

$$\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big[\exp\Big\{\frac{1}{\zeta}\, y \cdot (\mathcal{M}_s^t - \mathcal{M}_{s-1}^t)\Big\}\Big| \mathcal{F}_{s-1}\Big] \;\le\; \exp\Big\{\frac{1}{2\zeta^2}\, \nu_0 \sigma^2 (w_s^t)^2\, y^{\mathsf{T}} X_s X_s^{\mathsf{T}} y\Big\} \;\le\; \exp\Big\{\frac{1}{2\zeta}\, w_s^t\, y^{\mathsf{T}} X_s X_s^{\mathsf{T}} y\Big\}.$$

Consequently we get

$$\mathbb{E}_{\boldsymbol{\theta}}^{\pi}[\mathcal{Z}_s^{y,t} | \mathcal{F}_{s-1}] \;\le\; \exp\left\{\frac{1}{\zeta}\Big(y \cdot \mathcal{M}_{s-1}^t - \frac{1}{2} y^{\mathsf{T}} \mathcal{J}_{s-1}^t y\Big)\right\} \;=\; \mathcal{Z}_{s-1}^{y,t}.$$

So $(\mathcal{Z}_s^{y,t}, \mathcal{F}_s)$ is a supermartingale for any given $y \in \mathbb{R}^2$ and $t = 1, 2, \ldots$

To derive inequality (3.13), recall equation (3.10), which states that $\mathcal{J}_t^t = \mathfrak{X}\mathcal{I}_t^t$ where $\mathfrak{X} = \left[\begin{smallmatrix} 2 & x_1+x_2 \\ x_1+x_2 & x_1^2+x_2^2 \end{smallmatrix}\right]$ and $\mathcal{I}_t^t = \frac{1}{2}\sum_{s=1}^t w_s^t \ge 0$. Let $\mathcal{V} \subset \mathbb{R}^2$ be the set of eigenvectors of $\mathfrak{X}$, and consider the eigendecomposition of $\mathfrak{X}$:

$$\mathfrak{X} = P \Lambda P^{\mathsf{T}},$$

where P is an orthogonal matrix that has the eigenvectors of $\mathfrak{X}$ in its columns, and Λ is a diagonal matrix that has the eigenvalues of $\mathfrak{X}$ in its diagonal entries. For $z > 0$, define a set of four vectors $\mathcal{S}_z := \{\pm\frac{1}{\sqrt{2}} z v : v \in \mathcal{V}\} = \{\pm\frac{1}{\sqrt{2}} z P_i : P_i \text{ is the } i^{\text{th}} \text{ column of } P\}$. Then we have

$$\begin{aligned} \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\big\|\big(\mathcal{J}_t^t\big)^{-1}\mathcal{M}_t^t\big\| > z,\; \mathcal{I}_t^t > \gamma\big\} \;&=\; \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{(\mathcal{M}_t^t)^{\mathsf{T}}(\mathcal{J}_t^t)^{-2}\mathcal{M}_t^t > z^2,\; \mathcal{I}_t^t > \gamma\big\} \\ &=\; \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{(\mathcal{M}_t^t)^{\mathsf{T}} P \Lambda^{-2} P^{\mathsf{T}} \mathcal{M}_t^t > z^2 (\mathcal{I}_t^t)^2,\; \mathcal{I}_t^t > \gamma\big\}. \end{aligned}$$

Letting $\psi := \Lambda^{-1} P^{\mathsf{T}} \mathcal{M}_t^t$ we deduce that

$$\begin{aligned} \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\big\|\big(\mathcal{J}_t^t\big)^{-1}\mathcal{M}_t^t\big\| > z,\; \mathcal{I}_t^t > \gamma\big\} \;&\le\; \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\|\psi\| > z\mathcal{I}_t^t,\; \mathcal{I}_t^t > \gamma\big\} \\ &\le\; \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\bigcup_{i=1,2}\Big\{|\psi_i| > \tfrac{1}{\sqrt{2}} z\mathcal{I}_t^t,\; \mathcal{I}_t^t > \gamma\Big\}\Big\}. \end{aligned}$$

Note that $|\psi_i| > \frac{1}{\sqrt{2}} z \mathcal{I}_t^t$ implies that $P_i^\mathsf{T} \mathcal{M}_t^t$, the i^{th} component of $P^\mathsf{T} \mathcal{M}_t^t$, has an absolute value larger than $\frac{1}{\sqrt{2}} z \lambda_i \mathcal{I}_t^t$, where $\lambda_i = P_i^\mathsf{T} \mathfrak{X} P_i$ is the i^{th} diagonal entry of Λ. Therefore, viewing $\mathcal{V}$ as a basis for $\mathbb{R}^2$, we have

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\big\|\big(\mathcal{J}_t^t\big)^{-1}\mathcal{M}_t^t\big\| > z,\ \mathcal{I}_t^t > \gamma\Big\} &\le \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\bigcup_{i=1,2}\Big\{|P_i^\mathsf{T}\mathcal{M}_t^t| > \tfrac{1}{\sqrt{2}} z \lambda_i \mathcal{I}_t^t,\ \mathcal{I}_t^t > \gamma\Big\}\Big\} \\
&\overset{\text{(a)}}{=} \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\bigcup_{i=1,2}\Big\{|P_i^\mathsf{T}\mathcal{M}_t^t| > \tfrac{1}{\sqrt{2}} z P_i^\mathsf{T} \mathcal{J}_t^t P_i,\ \mathcal{I}_t^t > \gamma\Big\}\Big\} \\
&= \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\bigcup_{i=1,2}\Big\{\Big|\big(\tfrac{1}{\sqrt{2}} z P_i\big)^\mathsf{T}\mathcal{M}_t^t\Big| > \big(\tfrac{1}{\sqrt{2}} z P_i\big)^\mathsf{T} \mathcal{J}_t^t \big(\tfrac{1}{\sqrt{2}} z P_i\big),\ \mathcal{I}_t^t > \gamma\Big\}\Big\} \\
&\overset{\text{(b)}}{=} \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\bigcup_{w\in\mathcal{S}_z}\big\{w\cdot\mathcal{M}_t^t > w^\mathsf{T}\mathcal{J}_t^t w,\ \mathcal{I}_t^t > \gamma\big\}\Big\} \\
&\overset{\text{(c)}}{\le} \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\bigcup_{w\in\mathcal{S}_z}\big\{\mathcal{Z}_t^{w,t} \ge e^{\frac{1}{2\zeta} w^\mathsf{T}\mathcal{J}_t^t w},\ \mathcal{I}_t^t > \gamma\big\}\Big\} \\
&\overset{\text{(d)}}{\le} \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\bigcup_{w\in\mathcal{S}_z}\big\{\mathcal{Z}_t^{w,t} \ge e^{\rho_0 z^2 \mathcal{I}_t^t},\ \mathcal{I}_t^t > \gamma\big\}\Big\} \\
&\le \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\bigcup_{w\in\mathcal{S}_z}\big\{\mathcal{Z}_t^{w,t} \ge e^{\rho_0 z^2 \gamma}\big\}\Big\},
\end{aligned}$$

where: $\rho_0 = (\lambda_1 \wedge \lambda_2)/(4\zeta)$, (a) and (d) follow because $\lambda_i = P_i^\mathsf{T} \mathfrak{X} P_i$ and $\mathcal{J}_t^t = \mathfrak{X}\mathcal{I}_t^t$, (b) follows by the definition of $\mathcal{S}_z = \{\pm\frac{1}{\sqrt{2}} z P_i : i = 1, 2\}$, and (c) follows by the definition of $\mathcal{Z}_t^{w,t}$ in (B.11). We therefore have

$$\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\big\|\big(\mathcal{J}_t^t\big)^{-1}\mathcal{M}_t^t\big\| > z,\ \mathcal{I}_t^t > \gamma\Big\} \overset{\text{(e)}}{\le} \sum_{w\in\mathcal{S}_z} \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\mathcal{Z}_t^{w,t} \ge e^{\rho_0 z^2 \gamma}\big\} \overset{\text{(f)}}{\le} \sum_{w\in\mathcal{S}_z} e^{-\rho_0 z^2 \gamma} \overset{\text{(g)}}{=} 4e^{-\rho_0 z^2 \gamma},$$

where: (e) follows by the union bound, (f) follows by the Markov's inequality and the fact that $(\mathcal{Z}_s^{w,t}, \mathcal{F}_s)$ is a supermartingale, and (g) follows because the cardinality of $\mathcal{S}_z$ is 4. We conclude the proof by letting $\rho = (1 \vee z)\rho_0 = (\lambda_1 \wedge \lambda_2)/\big(4\big(\nu_0 \vee (x^*/x_0)\big)\sigma^2\big)$. Q.E.D.

Proof of Theorem 2. In the context of Section 3, the loss of a policy stems from three sources: (i) estimation inaccuracy due to changes in demand parameters, (ii) estimation errors due to noise, and (iii) price experimentation. To separate the effect of (iii) from (i-ii), note that

$$\begin{aligned}
\Delta_{\boldsymbol{\theta}}^{\pi}(T) &= \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{\sum_{t=1}^{T}\Big(1 - \frac{r(p_t,\theta_t)}{r^*(\theta_t)}\Big)\Big\} \\
&= \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{\sum_{t=1}^{T}\Big(1 - \frac{r(p_t,\theta_t)}{r^*(\theta_t)}\Big)\mathbb{I}\{t \in \mathcal{X}\}\Big\} + \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{\sum_{t=1}^{T}\Big(1 - \frac{r(p_t,\theta_t)}{r^*(\theta_t)}\Big)\mathbb{I}\{t \notin \mathcal{X}\}\Big\} \\
&\le 3n^{-1}T + \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{\sum_{t=1}^{T}\Big(1 - \frac{r(p_t,\theta_t)}{r^*(\theta_t)}\Big)\mathbb{I}\{t \notin \mathcal{X}\}\Big\} \qquad \text{(B.12)}
\end{aligned}$$

because the cardinality of $\mathcal{X}$ is less than or equal to $2(T/n + 1) \le 3T/n$. The expected sum on the right hand side above is bounded above as follows:

$$\begin{aligned}
&\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{\sum_{t=1}^{T}\Big(1-\frac{r(p_t,\theta_t)}{r^*(\theta_t)}\Big)\mathbb{I}\{t\not\in\mathcal{X}\}\Big\} \\
&\qquad\qquad = \ \mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{\sum_{t=0}^{T-1}\Big(1-\frac{r(p_{t+1},\theta_{t+1})}{r^*(\theta_{t+1})}\Big)\mathbb{I}\{t+1\not\in\mathcal{X}\}\Big\} \\
&\qquad\qquad = \ n^2 \ - \ \frac{\beta_{t+1}}{r^*(\theta_{t+1})}\sum_{t=n^2}^{T-1}\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{\big(\varphi(\theta_{t+1})-p_{t+1}\big)^2\,\mathbb{I}\{t+1\not\in\mathcal{X}\}\Big\} \\
&\qquad\qquad \le \ n^2 \ + \ c_2\sum_{t=n^2}^{T-1}\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{\big(\varphi(\theta_{t+1})-p_{t+1}\big)^2\,\mathbb{I}\{t+1\not\in\mathcal{X}\}\Big\} \\
&\qquad\qquad \overset{\text{(a)}}{\le} \ n^2 \ + \ 2K_0c_2\sum_{t=n^2}^{T-1}\Big(\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\|(\mathcal{J}^t_t)^{-1}\mathcal{W}^t_t\|^2+\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\|(\mathcal{J}^t_t)^{-1}\mathcal{M}^t_t\|^2\Big), \qquad\qquad \text{(B.13)}
\end{aligned}$$

where $c_2=\max_{(\alpha,\beta)\in\Theta}\{4\beta^2/\alpha^2\}$, $K_0=\max_{j=1,2}\big\{\max_{\theta}\{(\partial\varphi(\theta)/\partial\theta_j)^2\}\big\}$, and (a) follows by invoking identity (3.4) for the price experimentation scheme (3.6-3.7). To characterize the magnitude of $\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\|(\mathcal{J}^t_t)^{-1}\mathcal{M}^t_t\|^2$, we use Lemma 2 as follows: if $\pi=M(\kappa,x_1,x_2)$, we have

$$\begin{aligned}
\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\big\|\big(\mathcal{J}^t_t\big)^{-1}\mathcal{M}^t_t\big\|^2 &= \int_0^{\infty}\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\Big\{\big\|\big(\mathcal{J}^t_t\big)^{-1}\mathcal{M}^t_t\big\|^2>\xi\Big\}\,d\xi\,, \\
&\overset{\text{(b)}}{=} \int_0^{\infty}\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\Big\{\big\|\big(\mathcal{J}^t_t\big)^{-1}\mathcal{M}^t_t\big\|^2>\xi,\,\mathcal{I}^t_t\ge n\Big\}\,d\xi\,, \\
&\overset{\text{(c)}}{\le} 4\int_0^{\infty} e^{-\rho(\sqrt{\xi}\wedge\xi)n}\,d\xi \\
&= 4\int_0^{1} e^{-\rho n\xi}\,d\xi+4\int_1^{\infty} e^{-\rho n\sqrt{\xi}}\,d\xi \\
&\le 12/(\rho n), \qquad\qquad \text{(B.14)}
\end{aligned}$$

for all $t\ge n^2$, where: (b) follows because $\mathcal{I}^t_t\ge n$ for all $t\ge n^2$ under $M(\kappa,x_1,x_2)$, and (c) follows by Lemma 2. Similarly, if $\pi=W(\mu,\kappa,x_1,x_2)$, we have $\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\|(\mathcal{J}^t_t)^{-1}\mathcal{M}^t_t\|^2\le 12/(\rho c_\mu n)$ for all $t\ge n^2$, because $\mathcal{I}^t_t\ge c_\mu n$ for all $t\ge n^2$ under $W(\mu,\kappa,x_1,x_2)$. Letting $\tilde{c}=1\wedge c_\mu$, we therefore get $\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\|(\mathcal{J}^t_t)^{-1}\mathcal{M}^t_t\|^2\le 12/(\rho\tilde{c}n)$ for all $t\ge n^2$. Using this inequality on the right hand side of (B.13) we get

$$\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{\sum_{t=1}^{T}\Big(1-\frac{r(p_t,\theta_t)}{r^*(\theta_t)}\Big)\mathbb{I}\{t\not\in\mathcal{X}\}\Big\} \ \le \ n^2+2K_0c_2\sum_{t=n^2}^{T-1}\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\|(\mathcal{J}^t_t)^{-1}\mathcal{W}^t_t\|^2+\frac{24K_0c_2T}{\rho\tilde{c}n}. \qquad \text{(B.15)}$$

By Lemma 1 and inequality (B.15), we deduce that $\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\big\{\sum_{t=1}^{T}\big(1-r(p_t,\theta_t)/r^*(\theta_t)\big)\mathbb{I}\{t\not\in\mathcal{X}\}\big\}\le n^2+2K_0c_1c_2T^{2/3}+24K_0c_2T/(\rho\tilde{c}n)$. Plugging the value of $n=\lceil\kappa T^{1/3}\rceil$ in this upper bound, we get $\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\big\{\sum_{t=1}^{T}\big(1-r(p_t,\theta_t)/r^*(\theta_t)\big)\mathbb{I}\{t\not\in\mathcal{X}\}\big\}\le 4\kappa^2T^{2/3}+2K_0c_1c_2T^{2/3}+24K_0c_2T^{2/3}/(\rho\tilde{c}\kappa)$. Combining this inequality with (B.12) we conclude that $\Delta^{\pi}_{\boldsymbol{\theta}}(T)\le CT^{2/3}$ for all $\boldsymbol{\theta}\in\mathcal{V}(T,B)$, where $C=3/\kappa+4\kappa^2+2K_0c_1c_2+24K_0c_2/(\rho\tilde{c}\kappa)$. Q.E.D.

Remark 5. While choosing $\kappa=2$ is sufficient for first-order optimality, it is possible to obtain a tighter upper bound by minimizing the value of C with respect to κ. To do that, one would choose $\kappa=\arg\min_{\kappa\ge 2}\{3/\kappa+4\kappa^2+2K_0c_1c_2+24K_0c_2/(\rho\tilde{c}\kappa)\}$. It is also possible to further tailor the choice of κ to a particular environment by simulating the performance of $M(\kappa,x_1,x_2)$ and $W(\mu,\kappa,x_1,x_2)$ and estimating their regret in that environment.

Proof of Theorem 3. As in the proof of Theorem 2, we first isolate the loss due to price experimentation from other losses by noting that

$$\begin{aligned}\Delta^\pi_{\boldsymbol\theta}(T) &= \mathbb{E}^\pi_{\boldsymbol\theta}\Bigg\{\sum_{t=1}^{T}\bigg(1-\frac{r(p_t,\theta_t)}{r^*(\theta_t)}\bigg)\Bigg\}\\ &= \mathbb{E}^\pi_{\boldsymbol\theta}\Bigg\{\sum_{t=1}^{T}\bigg(1-\frac{r(p_t,\theta_t)}{r^*(\theta_t)}\bigg)\mathbb{I}\{t\in\mathcal{X}\}\Bigg\} + \mathbb{E}^\pi_{\boldsymbol\theta}\Bigg\{\sum_{t=1}^{T}\bigg(1-\frac{r(p_t,\theta_t)}{r^*(\theta_t)}\bigg)\mathbb{I}\{t\notin\mathcal{X}\}\Bigg\}\\ &\le 2n\bigg\lceil\frac{T}{n^2}\bigg\rceil + \mathbb{E}^\pi_{\boldsymbol\theta}\Bigg\{\sum_{t=2n+1}^{T}\bigg(1-\frac{r(p_t,\theta_t)}{r^*(\theta_t)}\bigg)\mathbb{I}\{t\notin\mathcal{X}\}\Bigg\}\\ &\le 4\kappa^{-1}B^{1/3}T^{2/3} + \mathbb{E}^\pi_{\boldsymbol\theta}\Bigg\{\sum_{t=2n+1}^{T}\bigg(1-\frac{r(p_t,\theta_t)}{r^*(\theta_t)}\bigg)\mathbb{I}\{t\notin\mathcal{X}\}\Bigg\}, \end{aligned}\tag{B.16}$$

because $t\in\mathcal{X}$ for all $t\le 2n$, and $n=\lceil\kappa B^{-1/3}T^{1/3}\rceil$. To find an upper bound on the expected sum on the right hand side above, we further note that

$$\begin{aligned}\mathbb{E}^\pi_{\boldsymbol\theta}\Bigg\{\sum_{t=2n+1}^{T}\bigg(1-\frac{r(p_t,\theta_t)}{r^*(\theta_t)}\bigg)\mathbb{I}\{t\notin\mathcal{X}\}\Bigg\} &= \mathbb{E}^\pi_{\boldsymbol\theta}\Bigg\{\sum_{t=2n}^{T-1}\bigg(1-\frac{r(p_{t+1},\theta_{t+1})}{r^*(\theta_{t+1})}\bigg)\mathbb{I}\{t+1\notin\mathcal{X}\}\Bigg\}\\ &= -\frac{\beta_{t+1}}{r^*(\theta_{t+1})}\sum_{t=2n}^{T-1}\mathbb{E}^\pi_{\boldsymbol\theta}\Big\{\big(\varphi(\theta_{t+1})-p_{t+1}\big)^2\,\mathbb{I}\{t+1\notin\mathcal{X}\}\Big\}\\ &\le c_2\sum_{t=2n}^{T-1}\mathbb{E}^\pi_{\boldsymbol\theta}\Big\{\big(\varphi(\theta_{t+1})-p_{t+1}\big)^2\,\mathbb{I}\{t+1\notin\mathcal{X}\}\Big\}, \end{aligned}\tag{B.17}$$

where $c_2=\max_{(\alpha,\beta)\in\Theta}\{4\beta^2/\alpha^2\}$. Here (3.4) and (3.15) imply that

$$\begin{aligned}\mathbb{E}^\pi_{\boldsymbol\theta}\Big\{\big(\varphi(\theta_{t+1})-p_{t+1}\big)^2\,\mathbb{I}\{t+1\notin\mathcal{X}\}\Big\} &\le \mathbb{E}^\pi_{\boldsymbol\theta}\big(\varphi(\theta_{t+1})-\varphi(\vartheta_{t+1})\big)^2\\ &\le 2K_0\,\mathbb{E}^\pi_{\boldsymbol\theta}\big\|\big(\mathcal{J}^t_t\big)^{-1}\mathcal{W}^t_t\big\|^2 + 2K_0\,\mathbb{E}^\pi_{\boldsymbol\theta}\big\|\big(\mathcal{J}^t_t\big)^{-1}\mathcal{M}^t_t\big\|^2, \end{aligned}\tag{B.18}$$

for all $t\ge 2n$, where $\vartheta_t=\arg\min_{\vartheta\in\Theta}\{\|\vartheta-\hat\theta_t\|\}$ is the truncated least squares estimate of θ_t, and $K_0=\max_{j=1,2}\big\{\max_\theta\{(\partial\varphi(\theta)/\partial\theta_j)^2\}\big\}$. To find an upper bound on the second term on the right hand side of inequality (B.18), we use the following result whose proof is identical to that of Lemma 2.

Lemma B.1. (exponential decay of estimation error due to noise) *Let π be $M_B(\kappa,x_1,x_2)$. Then there exists a finite positive constant ρ such that $\mathbb{P}^\pi_{\boldsymbol\theta}\{\|(\mathcal{J}^t_t)^{-1}\mathcal{M}^t_t\|>z,\ \mathcal{I}^t_t>\gamma\}\le 4\,e^{-\rho(z\wedge z^2)\gamma}$ for all $\boldsymbol\theta=(\theta_1,\dots,\theta_T)$, $z>0$, $\gamma>0$, and $t\ge 2$.*

Replacing Lemma 2 with Lemma B.1 in the argument used to derive (B.14), and using the fact that $\mathcal{I}^t_t\ge n$ for all $t\ge 2n$ under $M_B(\kappa,x_1,x_2)$, we deduce that $\mathbb{E}^\pi_{\boldsymbol\theta}\big\|\big(\mathcal{J}^t_t\big)^{-1}\mathcal{M}^t_t\big\|^2\le 12/(\rho n)$ for all $t\ge 2n$ under $M_B(\kappa,x_1,x_2)$. Thus, inequality (B.18) becomes

$$\mathbb{E}^\pi_{\boldsymbol\theta}\Big\{\big(\varphi(\theta_{t+1})-p_{t+1}\big)^2\,\mathbb{I}\{t+1\notin\mathcal{X}\}\Big\} \le 2K_0\,\mathbb{E}^\pi_{\boldsymbol\theta}\big\|\big(\mathcal{J}^t_t\big)^{-1}\mathcal{W}^t_t\big\|^2 + \frac{24K_0}{\rho n}, \tag{B.19}$$

for all $t\ge n$. Summing over $t=2n,\dots,T-1$, we obtain

$$\sum_{t=2n}^{T-1}\mathbb{E}^\pi_{\boldsymbol\theta}\Big\{\big(\varphi(\theta_{t+1})-p_{t+1}\big)^2\,\mathbb{I}\{t+1\notin\mathcal{X}\}\Big\} \le 2K_0\sum_{t=2n}^{T-1}\mathbb{E}^\pi_{\boldsymbol\theta}\big\|\big(\mathcal{J}^t_t\big)^{-1}\mathcal{W}^t_t\big\|^2 + \frac{24K_0}{\rho}n^{-1}T. \tag{B.20}$$

Next we use the following counterpart of Lemma 1, whose proof is at the end of this section.

Lemma B.2. (upper bound on aggregate estimation inaccuracy) *There exists a finite positive constant c_1, such that under $M_B(\kappa, x_1, x_2)$, $\sum_{t=2n}^{T-1} \|(\mathcal{J}_t^t)^{-1}\mathcal{W}_t^t\|^2 \leq c_1 B^{1/3}T^{2/3}$ almost surely for all $T = 1, 2, \ldots$ and $\boldsymbol{\theta} \in \mathcal{V}(T, B)$.*

Invoking Lemma B.2, and recalling that $n \geq \kappa B^{-1/3}T^{1/3}$, we deduce that

$$\sum_{t=2n}^{T-1} \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{ \big(\varphi(\theta_{t+1}) - p_{t+1}\big)^2 \, \mathbb{I}\{t+1 \notin \mathcal{X}\} \Big\} \;\leq\; 2K_0 c_1 B^{1/3}T^{2/3} + \frac{24K_0}{\rho\kappa}\, B^{1/3}T^{2/3} \;=\; c_3 B^{1/3}T^{2/3},$$

where $c_3 = 2K_0c_1 + 24K_0/(\rho\kappa)$. Combining the preceding inequality with inequalities (B.16-B.17), we conclude that $\Delta_{\boldsymbol{\theta}}^{\pi}(T) \leq CB^{1/3}T^{2/3}$ for all $\boldsymbol{\theta} \in \mathcal{V}(T, B)$, where $C = 4/\kappa + c_2c_3$. Q.E.D.

Proof of Lemma B.2. By the arguments used to derive (B.4), we deduce that $\sum_{t=2n}^{T-1} \big\|\big(\mathcal{J}_t^t\big)^{-1}\mathcal{W}_t^t\big\|^2 \leq 4n^2 V_{\boldsymbol{\theta}}(T)$ under $M_B(\kappa, x_1, x_2)$. Combining this inequality with (2.5), we conclude that

$$\sum_{t=2n}^{T-1} \big\|\big(\mathcal{J}_t^t\big)^{-1}\mathcal{W}_t^t\big\|^2 \;\leq\; 4n^2 B \;\overset{(a)}{\leq}\; 16\kappa^2 B^{1/3}T^{2/3}, \tag{B.21}$$

where (a) follows because $n = \lceil \kappa B^{-1/3}T^{1/3} \rceil$. We get the desired result by letting $c_1 = 16\kappa^2$. Q.E.D.

Appendix C: Proof of the results in Section 4. Note that if $\tau_j^* \geq \tau_{j+1}^* - 2$, then the expected value on the right hand side of (4.11) is less than $2n \leq 4\kappa\sqrt{T}$, which gives us an upper bound on the losses due to detection and estimation between j^{th} and $(j+1)^{\text{st}}$ change-points. Therefore, we hereafter focus on the case $\tau_j^* < \tau_{j+1}^* - 2$ for any given $j = 0, \ldots, \mathcal{C}$.

Proof of Lemma 3. Recall that the ϵ_t have a light-tailed distribution, which implies that there exist finite constants x_0 and ν_0 such that $\mathbb{E}_{\boldsymbol{\theta}}^{\pi}[\exp(x\epsilon_t)] \leq \exp(\frac{1}{2}\nu_0\sigma^2x^2)$ for all x satisfying $|x| \leq x_0$. Choosing $c_\epsilon = (6/x_0) \vee (12\nu_0\sigma^2)$, we prove (4.12) for $\pi = D(\eta, \kappa, x_1, x_2)$ with $\kappa = c_\epsilon/(\eta \wedge \eta^2)$. Note that, for the case $\epsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, we have $c_\epsilon = 12\sigma^2$.

Given a real number y with $|y| = \frac{1}{2}\eta$, let $\{Z_t^y, t = 1, 2, \ldots\}$ be a stochastic process satisfying $Z_0^y = 1$ and

$$Z_t^y := \exp\big\{ \tfrac{1}{\zeta}\big(yS_t - \tfrac{1}{2}y^2t\big) \big\} \qquad \text{for all } t = 1, 2, \ldots \tag{C.1}$$

where $\zeta = \big(\frac{\eta}{2x_0}\big) \vee (\nu_0\sigma^2)$ and $S_t = \sum_{q=1}^{t} \epsilon_q$. First note that Z_t^y is integrable for all t. Let $\mathcal{F}_t := \sigma(\epsilon_1, \ldots, \epsilon_t)$. Then, we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}^{\pi}[Z_t^y | \mathcal{F}_{t-1}] \;&=\; \exp\big\{ \tfrac{1}{\zeta}\big(yS_{t-1} - \tfrac{1}{2}y^2t\big) \big\}\, \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\big[\exp\big(\tfrac{1}{\zeta}\, y\epsilon_t\big) | \mathcal{F}_{t-1}\big] \\
&\overset{(a)}{\leq}\; \exp\big\{ \tfrac{1}{\zeta}\big(yS_{t-1} - \tfrac{1}{2}y^2t\big) \big\} \exp\big\{ \tfrac{1}{2\zeta^2}\, \nu_0\sigma^2y^2 \big\} \\
&\overset{(b)}{\leq}\; \exp\big\{ \tfrac{1}{\zeta}\big(yS_{t-1} - \tfrac{1}{2}y^2(t-1)\big) \big\} \\
&=\; Z_{t-1}^y.
\end{aligned} \tag{C.2}$$

for all $t = 1, 2, \ldots$, where: (a) follows because $|y/\zeta| = |\eta/(2\zeta)| \leq x_0$, and (b) follows because $\nu_0\sigma^2 \leq \zeta$. Thus $(Z_t^y, \mathcal{F}_t)$ is a supermartingale for all $y \in \mathbb{R}$. Now note that

$$\bar{\epsilon}_{10} \;=\; \frac{1}{m}\sum_{t \in \mathcal{X}_{10}} \epsilon_t \;=\; \frac{1}{m}\sum_{t=1}^{m} \epsilon_t \;=\; \frac{S_m}{m}. \tag{C.3}$$

Therefore, we have

$$\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{ \big|\bar{\epsilon}_{10}\big| \geq \tfrac{1}{2}\eta \big\} \;=\; \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{ \big|S_m\big| \geq \tfrac{1}{2}\eta m \big\}. \tag{C.4}$$

Choosing $y = \frac{1}{2}\eta$, we get

$$\begin{aligned}
\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\{S_m \ge \tfrac{1}{2}\eta m\} &= \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{S_m \ge ym\big\} \\
&= \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{yS_m - \tfrac{1}{2}y^2 m \ge \tfrac{1}{2}y^2 m\big\} \\
&\le \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{Z^y_m \ge \exp\big(\tfrac{1}{2\zeta}y^2 m\big)\big\} \\
&\overset{(c)}{\le} \mathbb{E}^{\pi}_{\boldsymbol{\theta}}\{Z^y_m\}\,\exp\big(-\tfrac{1}{2\zeta}y^2 m\big) \\
&\overset{(d)}{\le} \exp\big(-\tfrac{1}{8\zeta}\eta^2 m\big) \\
&\overset{(e)}{\le} \exp\big(-\tfrac{1}{8\zeta}\eta^2 \kappa \log T\big) \\
&= T^{-\frac{1}{8\zeta}\eta^2\kappa} \\
&\overset{(f)}{\le} T^{-3/2},
\end{aligned} \tag{C.5}$$

where: (c) follows from Markov's inequality, (d) follows because $y^2 = \frac{1}{4}\eta^2$ and Z^y_t is a supermartingale with $Z^y_0 = 1$, (e) follows because $m = \lceil \kappa \log T\rceil \ge \kappa \log T$, and (f) follows because $\kappa \ge \frac{6}{\eta x_0} \vee \frac{12\nu_0\sigma^2}{\eta^2} = 12\zeta\eta^{-2}$. Similarly, choosing $y = -\frac{1}{2}\eta$, we deduce by the argument used for deriving (C.5) that $\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\{S_m \le -\frac{1}{2}\eta m\} \le T^{-3/2}$. Therefore, $\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\big|\bar{\epsilon}_{10}\big| \ge \frac{1}{2}\eta\big\} = \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\big|S_m\big| \ge \frac{1}{2}\eta m\big\} \le 2T^{-3/2}$. Because the experimentation sets $\mathcal{X}_{ik}$ are disjoint and $\{\epsilon_t\}$ are independent and identically distributed random variables, $\bar{\epsilon}_{10}$ has the same distribution as $\bar{\epsilon}_{ik}$ for all i and k. Therefore, by the above argument, we have $\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\big|\bar{\epsilon}_{ik}\big| \ge \frac{1}{2}\eta\big\} \le 2T^{-3/2}$ for all i and k. Q.E.D.

Proof of Lemma 4. Define

$$A_j := \bigcup_{k=L(\tau^*_j)}^{\tau^*_j} \big\{\theta_s = \theta_t \ne \theta_{(\tau^*_j+1)n+1} \text{ for all } s,t \in \mathcal{X}_{1k} \cup \mathcal{X}_{2k}\big\}, \tag{C.6}$$

the event that there is at least one cycle k between $L(\tau^*_j)$ and τ^*_j such that there is no change-point in $\mathcal{X}_{1k} \cup \mathcal{X}_{2k}$, and the value of the demand parameter vector during the periods in $\mathcal{X}_{1k} \cup \mathcal{X}_{2k}$ is different than the one after the j^{th} change-point. We first calculate the loss due to detection delay on A_j. Let $\mathcal{D}_j = \hat{\tau}^+_j - \tau^*_j$ be the delay of the true detection following the j^{th} change-point. Assuming $\tau^*_j < \tau^*_{j+1} - 2$, we have

$$\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Bigg\{\sum_{s=n\tau^*_j+1}^{n\hat{\tau}^+_j}\bigg(1 - \frac{r(p_s,\theta_s)}{r^*(\theta_s)}\bigg)\mathbb{I}_{\{s \notin \mathcal{X}\}\cap A_j}\Bigg\} \le n\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\big\{\mathcal{D}_j\,\mathbb{I}_{A_j}\big\}. \tag{C.7}$$

Note that $\mathcal{D}_0 = 0$ by definition, and for $j = 1, 2, \ldots, \mathcal{C}$, the expected value of $\mathcal{D}_j\,\mathbb{I}_{A_j}$ can be bounded above as follows:

$$\begin{aligned}
\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\big\{\mathcal{D}_j\,\mathbb{I}_{A_j}\big\} &= \sum_{d=1}^{\tau^*_{j+1}-\tau^*_j} \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\mathcal{D}_j \ge d\,,\, A_j\big\} \\
&\overset{(a)}{\le} 2 + \sum_{d=3}^{\tau^*_{j+1}-\tau^*_j} \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\mathcal{D}_j \ge d\,,\, A_j\big\} \\
&\overset{(b)}{\le} 2 + \sum_{d=3}^{\tau^*_{j+1}-\tau^*_j} \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\chi_{\tau^*_j+d-1} = 0\,,\, A_j\big\},
\end{aligned} \tag{C.8}$$

where: (a) follows because $\mathbb{P}^{\pi}_{\boldsymbol{\theta}}(\cdot)$ is a probability measure, and (b) follows because $\mathcal{D}_j \geq d \geq 3$ implies that there was no detection in cycle $\tau_j^* + d - 2$. On A_j, the probability of no detection in cycle $k^* = \tau_j^* + 1, \ldots, \tau_{j+1}^* - 2$ is

$$\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\chi_{k^*+1} = 0\,,\, A_j\big\} \;=\; \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\Big\{\sup\nolimits_{i,k}\big\{|\bar{D}_{ik^*} - \bar{D}_{ik}| : L(\tau_j^*) \leq k < k^*\big\} \leq \eta\,,\, A_j\Big\}. \tag{C.9}$$

Note that, on the event A_j, there exists a cycle $k_0 = L(\tau_j^*), \ldots, \tau_j^*$ such that for all $s, t \in \mathcal{X}_{1k_0} \cup \mathcal{X}_{2k_0}$ we have $\theta_s = \theta_t \neq \theta_{n(\tau_j^*+1)+1}$. Let $y_0 := \theta_{nk_0+1}$ and $y^* := \theta_{n(\tau_j^*+1)+1}$. Then, by the preceding identity, we have the following for $k^* = \tau_j^* + 1, \ldots, \tau_{j+1}^* - 2$:

$$\begin{aligned}
\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\chi_{k^*+1} = 0\,,\, A_j\big\} \;&\leq\; \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\big|\bar{D}_{ik^*} - \bar{D}_{ik_0}\big| \leq \eta \text{ for } i = 1,2\,,\, A_j\big\} \\
&=\; \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\left\{\frac{1}{m}\left|\sum_{t\in\mathcal{X}_{ik^*}} D_t \;-\; \sum_{s\in\mathcal{X}_{ik_0}} D_s\right| \leq \eta \text{ for } i = 1,2\,,\, A_j\right\} \\
&\overset{\text{(c)}}{=}\; \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\left\{\frac{1}{m}\left|\sum_{t\in\mathcal{X}_{ik^*}} (\tilde{X}_i\cdot\theta_t + \epsilon_t) \;-\; \sum_{s\in\mathcal{X}_{ik_0}} (\tilde{X}_i\cdot\theta_s + \epsilon_s)\right| \leq \eta \text{ for } i = 1,2\,,\, A_j\right\} \\
&\overset{\text{(d)}}{=}\; \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\left\{\frac{1}{m}\left|\sum_{t\in\mathcal{X}_{ik^*}} (\tilde{X}_i\cdot y^* + \epsilon_t) \;-\; \sum_{s\in\mathcal{X}_{ik_0}} (\tilde{X}_i\cdot y_0 + \epsilon_s)\right| \leq \eta \text{ for } i = 1,2\,,\, A_j\right\} \\
&=\; \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\left\{\frac{1}{m}\left|m\,\tilde{X}_i\cdot(y_0 - y^*) + \sum_{t\in\mathcal{X}_{ik^*}} \epsilon_t \;-\; \sum_{s\in\mathcal{X}_{ik_0}} \epsilon_s\right| \leq \eta \text{ for } i = 1,2\,,\, A_j\right\} \\
&\overset{\text{(e)}}{\leq}\; \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\left\{\frac{1}{m}\left|\sum_{t\in\mathcal{X}_{ik^*}} \epsilon_t \;-\; \sum_{s\in\mathcal{X}_{ik_0}} \epsilon_s\right| \geq \big|\tilde{X}_i\cdot(y_0 - y^*)\big| - \eta \text{ for } i = 1,2\,,\, A_j\right\} \\
&=\; \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\Big\{\big|\bar{\epsilon}_{ik^*} - \bar{\epsilon}_{ik_0}\big| \geq \big|\tilde{X}_i\cdot(y_0 - y^*)\big| - \eta \text{ for } i = 1,2\,,\, A_j\Big\},
\end{aligned} \tag{C.10}$$

where: $\tilde{X}_i := \left[\begin{smallmatrix}1\\ x_i\end{smallmatrix}\right]$ and $\bar{\epsilon}_{ik} = m^{-1}\sum_{t\in\mathcal{X}_{ik}} \epsilon_t$ for all i, k, (c) follows by (3.3) and the fact that $p_t = x_i$ for all $t \in \mathcal{X}_{ik}$, (d) follows because $\theta_t = y^*$ for all $t \in \mathcal{X}_{ik^*}$ and $\theta_s = y_0$ for all $s \in \mathcal{X}_{ik_0}$, and (e) follows by triangle inequality. Because $y_0 \neq y^*$, we know by condition (4.1) that $\|y_0 - y^*\| \geq \delta$. By elementary algebra, this implies that $\big|\tilde{X}_i\cdot(y_0 - y^*)\big| \geq a\delta$ for some $i_0 = 1,2$, where $a = \frac{|x_1 - x_2|}{4(1\vee x_1\vee x_2)}$. Recalling that $\eta = \frac{|x_1 - x_2|}{8(1\vee x_1\vee x_2)}\,\delta = \frac{1}{2}a\delta$, we have

$$\begin{aligned}
\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\chi_{k^*+1} = 0\,,\, A_j\big\} \;&\leq\; \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\big|\bar{\epsilon}_{i_0k^*} - \bar{\epsilon}_{i_0k_0}\big| \geq \tfrac{1}{2}a\delta\,,\, A_j\big\} \\
&=\; \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\big|\bar{\epsilon}_{i_0k^*} - \bar{\epsilon}_{i_0k_0}\big| \geq \eta\,,\, A_j\big\} \\
&\leq\; \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\big|\bar{\epsilon}_{i_0k^*}\big| \geq \tfrac{1}{2}\eta\,,\, A_j\big\} + \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\big|\bar{\epsilon}_{i_0k_0}\big| \geq \tfrac{1}{2}\eta\,,\, A_j\big\} \\
&\overset{\text{(f)}}{\leq}\; 2\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\Big\{\bigcup\nolimits_{i,k}\big\{\big|\bar{\epsilon}_{ik}\big| \geq \tfrac{1}{2}\eta\big\}\,,\, A_j\Big\} \\
&\leq\; 2\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\Big\{\bigcup\nolimits_{i,k}\big\{\big|\bar{\epsilon}_{ik}\big| \geq \tfrac{1}{2}\eta\big\}\Big\} \\
&\overset{\text{(g)}}{\leq}\; 2\sum\nolimits_{i,k}\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\big|\bar{\epsilon}_{ik}\big| \geq \tfrac{1}{2}\eta\big\},
\end{aligned} \tag{C.11}$$

for $k^* = \tau_j^* + 1, \ldots, \tau_{j+1}^* - 2$, where: (f) follows because for any given i and k, $\big|\bar{\epsilon}_{ik}\big| \geq \frac{1}{4}a\delta$ implies $\bigcup_{i,k}\big\{\big|\bar{\epsilon}_{ik}\big| \geq \frac{1}{4}a\delta\big\}$, and (g) follows by the union bound. Using the bound in Lemma 3 on the right hand side of (C.11), we get

$$\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\chi_{k^*+1} = 0\,,\, A_j\big\} \;\leq\; 4\sum\nolimits_{i,k} T^{-3/2} \;\leq\; 8T^{-3/2}\lfloor T/n\rfloor \;\leq\; 8n^{-1}T^{-1/2}\,, \tag{C.12}$$

for $k^* = \tau_j^* + 1, \ldots, \tau_{j+1}^* - 2$. Combining (C.8) and (C.12), we deduce that

$$\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\{\mathcal{D}_j\,\mathbb{I}_{A_j}\} \;\le\; 2 + 8\sum_{d=3}^{\tau_{j+1}^* - \tau_j^*} n^{-1}T^{-1/2} \;\le\; 2 + 8n^{-1}T^{-1/2}\lfloor T/n \rfloor \;\le\; 2 + 8n^{-2}T^{1/2} \;\le\; 2 + 8\kappa^{-2}T^{-1/2}.$$

Recalling (C.7), we conclude that

$$\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\left\{\sum_{s=n\tau_j^*+1}^{n\hat{\tau}_j^+}\left(1 - \frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}_{\{s\notin\mathcal{X}\}\cap A_j}\right\} \;\le\; (2 + 8\kappa^{-2}T^{-1/2})n \;\le\; 4\kappa\sqrt{T} + 16\kappa^{-1}. \tag{C.13}$$

Now we find an upper bound on loss due to detection delay on the event A_j^c. Assuming $\tau_j^* < \tau_{j+1}^* - 2$, we have the following on A_j^c:

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\left\{\sum_{s=n\tau_j^*+1}^{n\hat{\tau}_j^+}\left(1 - \frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}_{\{s\notin\mathcal{X}\}\cap A_j^c}\right\} \\
&\qquad \le\; n + \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\left\{\sum_{s=n(\tau_j^*+1)+1}^{n\tau_{j+1}^*}\left(1 - \frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}_{\{s\notin\mathcal{X},\, s\le n\hat{\tau}_j^+\}\cap A_j^c}\right\} \\
&\qquad \overset{(a')}{\le}\; n + c_1\sum_{s=n(\tau_j^*+1)+1}^{n\tau_{j+1}^*}\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\left\{\big(\varphi(\theta_s) - p_s\big)^2\,\mathbb{I}_{\{s\notin\mathcal{X},\, s\le n\hat{\tau}_j^+\}\cap A_j^c}\right\} \\
&\qquad \le\; n + c_2\sum_{s=n(\tau_j^*+1)+1}^{n\tau_{j+1}^*}\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\left\{\|\theta_s - \hat{\theta}_s\|^2\,\mathbb{I}_{\{s\notin\mathcal{X},\, s\le n\hat{\tau}_j^+\}\cap A_j^c}\right\} \\
&\qquad =\; n + c_2\sum_{s=n(\tau_j^*+1)}^{n\tau_{j+1}^*-1}\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\left\{\|\theta_{s+1} - \hat{\theta}_{s+1}\|^2\,\mathbb{I}_{\{s+1\notin\mathcal{X},\, s< n\hat{\tau}_j^+\}\cap A_j^c}\right\},
\end{aligned} \tag{C.14}$$

where: $c_1 = \max_{(\alpha,\beta)\in\Theta}\{4\beta^2/\alpha^2\}$, $c_2 = c_1\max_{i=1,2}\big\{\max_\theta\{(\partial\varphi(\theta)/\partial\theta_i)^2\}\big\}$, and (a′) follows by definitions of $r(\cdot,\cdot)$, $r^*(\cdot)$, and $\varphi(\cdot)$. By (3.4), we have

$$\hat{\theta}_{s+1} - \theta_{s+1} \;=\; \big(\mathcal{J}_s^s\big)^{-1}\sum_{q=1}^{s} w_q^s X_q X_q^{\mathsf{T}}(\theta_q - \theta_{s+1}) \;+\; \big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s \quad \text{for all } s. \tag{C.15}$$

The preceding identity implies the following for $s = n(\tau_j^*+1), \ldots, n\tau_{j+1}^* - 1$ satisfying $s+1 \notin \mathcal{X}$:

$$\begin{aligned}
\|\hat{\theta}_{s+1} - \theta_{s+1}\| &\overset{(b')}{\le}\; \big(\mathcal{I}_s^s\big)^{-1}\sum_{q=1}^{s} w_q^s\,\|\theta_q - \theta_{s+1}\| \;+\; \big\|\big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s\big\| \\
&\overset{(c')}{\le}\; \big(\mathcal{I}_s^s\big)^{-1}\sum_{q=nL(\tau_j^*)+1}^{n(\tau_j^*+1)} w_q^s\,\|\theta_q - \theta_{s+1}\| \;+\; \big\|\big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s\big\|,
\end{aligned} \tag{C.16}$$

where: (b′) follows by triangle inequality and the fact that the eigenvalues of $\big(\mathcal{J}_s^s\big)^{-1}X_qX_q^{\mathsf{T}}$ are 0 and $\pm\big(\mathcal{I}_s^s\big)^{-1}$ for all $q \in \mathcal{X}$, and (c′) follows because $w_q^s = 0$ for $q \le nL(\tau_j^*) \le n(\tau_j^*+1) \le s$ and $\|\theta_q - \theta_{s+1}\| = 0$ for $(\tau_j^*+1)n + 1 \le q \le s \le \tau_{j+1}^* n - 1$. By the definition of A_j in (C.6) we have the following on A_j^c: for all cycles k between $L(\tau_j^*)$ and τ_j^*, either (i) there is a change-point in $\mathcal{X}_{1k}\cup\mathcal{X}_{2k}$, or (ii) the value of the demand parameter vector during the periods in $\mathcal{X}_{1k}\cup\mathcal{X}_{2k}$ is exactly the same as the one after the j^{th} change-point. Letting $\mathcal{K}_j^{(\text{i})}$ and $\mathcal{K}_j^{(\text{ii})}$ be the sets of cycles between $L(\tau_j^*)$ and τ_j^* that satisfy conditions (i) and (ii), respectively, we re-express (C.16) as follows:

$$\begin{aligned}
\|\hat{\theta}_{s+1}-\theta_{s+1}\| &\le \big(\mathcal{I}_s^s\big)^{-1}\Bigg(\sum_{k\in\mathcal{K}_j^{(\text{i})}}+\sum_{k\in\mathcal{K}_j^{(\text{ii})}}\Bigg)\sum_{q=nk+1}^{n(k+1)} w_q^s\,\|\theta_q-\theta_{s+1}\| \;+\; \big\|\big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s\big\| \\
&\overset{(\text{d}')}{\le} \big(\mathcal{I}_s^s\big)^{-1}\sum_{k\in\mathcal{K}_j^{(\text{i})}}\sum_{q=nk+1}^{n(k+1)} w_q^s\,\|\theta_q-\theta_{s+1}\| \;+\; \big\|\big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s\big\| \\
&\overset{(\text{e}')}{\le} \big(\mathcal{I}_s^s\big)^{-1}c_3\sum_{k\in\mathcal{K}_j^{(\text{i})}}\sum_{q=nk+1}^{n(k+1)} w_q^s \;+\; \big\|\big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s\big\| \\
&\overset{(\text{f}')}{\le} \big(\mathcal{I}_s^s\big)^{-1}c_3\mathcal{C}m \;+\; \big\|\big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s\big\|, \qquad\qquad (\text{C.17})
\end{aligned}$$

for $s=n(\tau_j^*+1),\ldots,n\tau_{j+1}^*-1$ satisfying $s+1\notin\mathcal{X}$, where: $c_3=\max_{\theta,\theta'\in\Theta}\|\theta-\theta'\|$, (d′) follows because given $k\in\mathcal{K}_j^{(\text{ii})}$, we have $w_q^s\,\|\theta_q-\theta_{s+1}\|=0$ for $q=nk+1,\ldots,n(k+1)$, (e′) follows because $\|\theta_q-\theta_{s+1}\|\le c_3$, and (f′) follows because $w_q^s=0$ if $s\notin\mathcal{X}$, and the cardinality of $\mathcal{K}_j^{(\text{i})}$ is less than or equal to the number of change-points, $\mathcal{C}$. Squaring and taking the expectation of both sides of (C.17), we get

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{\|\hat{\theta}_{s+1}-\theta_{s+1}\|^2\,\mathbb{I}_{\{s+1\notin\mathcal{X},\,s<n\hat{\tau}_j^+\}\cap A_j^c}\Big\} \\
&\qquad\le \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{\Big(2\big(\mathcal{I}_s^s\big)^{-2}c_3^2\mathcal{C}^2m^2+2\big\|\big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s\big\|^2\Big)\,\mathbb{I}_{\{s+1\notin\mathcal{X},\,s<n\hat{\tau}_j^+\}\cap A_j^c}\Big\}. \qquad (\text{C.18})
\end{aligned}$$

On A_j^c, we know that $\mathcal{I}_s^s=2\big(\lceil s/n\rceil-L(\tau_j^*)+1\big)m\ge 2\big(\lceil s/n\rceil-\tau_j^*+1\big)m$ for $s=n(\tau_j^*+1),\ldots,n\hat{\tau}_j^+-1$ satisfying $s+1\notin\mathcal{X}$. Thus, (C.18) implies that

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{\|\hat{\theta}_{s+1}-\theta_{s+1}\|^2\,\mathbb{I}_{\{s+1\notin\mathcal{X},\,s<n\hat{\tau}_j^+\}\cap A_j^c}\Big\} \\
&\qquad\le \frac{c_3^2\mathcal{C}^2}{2\big(\lceil s/n\rceil-\tau_j^*+1\big)^2} \;+\; 2\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{\big\|\big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s\big\|^2\,\mathbb{I}_{\{s+1\notin\mathcal{X},\,s<n\hat{\tau}_j^+\}\cap A_j^c}\Big\} \\
&\qquad\overset{(\text{g}')}{\le} \frac{c_3^2\mathcal{C}^2}{2\big(\lceil s/n\rceil-\tau_j^*+1\big)^2} \;+\; \frac{12}{\rho\big(\lceil s/n\rceil-\tau_j^*+1\big)m}, \qquad (\text{C.19})
\end{aligned}$$

where (g′) follows by the arguments used to prove inequality (B.14) and Lemma 2. Summing both sides of (C.19) over $s=n(\tau_j^*+1),\ldots,n\tau_{j+1}^*-1$, we deduce that

$$\begin{aligned}
&\sum_{s=n(\tau_j^*+1)}^{n\tau_{j+1}^*-1} \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{\|\theta_{s+1}-\hat{\theta}_{s+1}\|^2\,\mathbb{I}_{\{s+1\notin\mathcal{X},\,s\le\hat{\tau}_j^+ n\}\cap A_j^c}\Big\} \\
&\qquad\le \sum_{s=n(\tau_j^*+1)}^{n\tau_{j+1}^*-1}\left(\frac{c_3^2\,\mathcal{C}^2}{2\big(\lceil s/n\rceil-\tau_j^*+1\big)^2}+\frac{12}{\rho\big(\lceil s/n\rceil-\tau_j^*+1\big)m}\right) \\
&\qquad\overset{(\mathrm{h}')}{\le} n\sum_{q=2}^{\tau_{j+1}^*-\tau_j^*+1}\left(\frac{c_3^2\,\mathcal{C}^2}{2q^2}+\frac{12}{\rho q m}\right) \\
&\qquad\le n\left(\frac{c_3^2\,\mathcal{C}^2\,\pi^2}{12}+\frac{12}{\rho m}\log(\tau_{j+1}^*-\tau_j^*+1)\right) \\
&\qquad\overset{(\mathrm{i}')}{\le} n\left(\frac{c_3^2\,\mathcal{C}^2\,\pi^2}{12}+\frac{6}{\rho\kappa}\right) \\
&\qquad\overset{(\mathrm{j}')}{\le} \left(\frac{c_3^2\,\mathcal{C}^2\,\pi^2\kappa}{6}+\frac{12}{\rho}\right)\sqrt{T},
\end{aligned} \tag{C.20}$$

for $T\ge 3$, where: (h′) follows by expressing the time index as $s=(\tau_j^*+q-1)n+i$, (i′) follows because $m\ge\kappa\log T\ge 2\kappa\log(\tau_{j+1}^*-\tau_j^*+1)$ for $T\ge 3$, and (j′) follows because $n\le 2\kappa\sqrt{T}$. By inequalities (C.14) and (C.20), we deduce that $\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\big\{\sum_{s=n\tau_j^*+1}^{n\hat{\tau}_j^+}\big(1-r(p_s,\theta_s)/r^*(\theta_s)\big)\mathbb{I}_{\{s\notin\mathcal{X}\}\cap A_j^c}\big\}\le c_4\sqrt{T}$, where $c_4=2\kappa+c_2\big(c_3^2\mathcal{C}^2\pi^2\kappa/6+12/\rho\big)$. Combining this result with (C.13), we conclude that

$$\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\left\{\sum_{s=n\tau_j^*+1}^{n\hat{\tau}_j^+}\left(1-\frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}_{\{s\notin\mathcal{X}\}}\right\}\le C_1\sqrt{T}, \tag{C.21}$$

where $C_1=4(\kappa+4\kappa^{-1})\vee c_4$. Q.E.D.

Proof of Lemma 5. Assume that $\tau_j^*<\tau_{j+1}^*-2$, and that there exists at least one false detection between the j^{th} and $(j+1)^{\text{st}}$ change-points. Let $\mathcal{E}_j=\tau_{j+1}^*-\hat{\tau}_j^-$ be the earliness of the first false detection after the j^{th} change-point. Then, we have

$$\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\left\{\sum_{s=n\hat{\tau}_j^-+1}^{n\tau_{j+1}^*}\left(1-\frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}_{\{s\notin\mathcal{X}\}}\right\}\;\le\; n\big(1+\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\{\mathcal{E}_j\}\big). \tag{C.22}$$

For $j=0,1,\ldots,\mathcal{C}$, the expected earliness of false detections before the $(j+1)^{\text{st}}$ change-point is given by

$$\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\{\mathcal{E}_j\}\;=\;\sum_{\varepsilon=1}^{\tau_{j+1}^*-\tau_j^*-2}\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\{\mathcal{E}_j\ge\varepsilon\}\;=\;\sum_{\varepsilon=1}^{\tau_{j+1}^*-\tau_j^*-2}\;\sum_{q=\varepsilon}^{\tau_{j+1}^*-\tau_j^*-2}\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\{\mathcal{E}_j=q\}. \tag{C.23}$$

By definition, $\tau_j^*<\hat{\tau}_j^+<\hat{\tau}_j^-\le\tau_{j+1}^*$, i.e., if there is a false detection between the j^{th} and $(j+1)^{\text{st}}$ change-points then it must be preceded by the true detection after the j^{th} change-point. Therefore, for all $q=1,\ldots,\tau_{j+1}^*-\tau_j^*-2$, the event $\mathcal{E}_j=q$ implies that the true detection after the j^{th} change-point is between cycles τ_j^*+1 and τ_{j+1}^*-q-2, and that there is a false detection in cycle τ_{j+1}^*-q-1. More formally, $\{\mathcal{E}_j=q\}\subseteq\big\{L(\tau_{j+1}^*-q-1)\ge\tau_j^*+1\big\}\cap\big\{\chi_{\tau_{j+1}^*-q}=1\big\}$, where $L(k)=\max\{\tau\le k:$

$\chi_\tau = 1\}$ is the latest detection cycle that precedes cycle k. Letting $B_{jq} := \big\{L(\tau^*_{j+1} - q - 1) \geq \tau^*_j + 1\big\}$, we have the following by (C.23):

$$\mathbb{E}^\pi_{\boldsymbol{\theta}}\{\mathcal{E}_j\} \;\leq\; \sum_{\varepsilon=1}^{\tau^*_{j+1}-\tau^*_j-2} \; \sum_{q=\varepsilon}^{\tau^*_{j+1}-\tau^*_j-2} \mathbb{P}^\pi_{\boldsymbol{\theta}}\big\{\chi_{\tau^*_{j+1}-q} = 1\,,\, B_{jq}\big\}. \tag{C.24}$$

For $k^* = \tau^*_{j+1} - q - 1$, the definition of the detection test (4.6) implies

$$\begin{aligned}
\mathbb{P}^\pi_{\boldsymbol{\theta}}\big\{\chi_{k^*+1} = 1\,,\, B_{jq}\big\} &= \mathbb{P}^\pi_{\boldsymbol{\theta}}\Big\{\sup\nolimits_{i,k}\big\{\big|\bar{D}_{ik^*} - \bar{D}_{ik}\big| : L(k^*) \leq k < k^*\big\} > \eta\,,\, B_{jq}\Big\} \\
&\overset{\text{(a)}}{\leq} \mathbb{P}^\pi_{\boldsymbol{\theta}}\Big\{\sup\nolimits_{i,k}\big\{\big|\bar{D}_{ik^*} - \bar{D}_{ik}\big| : \tau^*_j + 1 \leq k < k^*\big\} > \eta\,,\, B_{jq}\Big\} \\
&= \mathbb{P}^\pi_{\boldsymbol{\theta}}\left\{\bigcup_{i=1}^{2} \; \bigcup_{k=\tau^*_j+1}^{k^*-1} \big\{\big|\bar{D}_{ik^*} - \bar{D}_{ik}\big| > \eta\big\}\,,\, B_{jq}\right\} \\
&\overset{\text{(b)}}{\leq} \sum_{i=1}^{2} \; \sum_{k=\tau^*_j+1}^{k^*-1} \mathbb{P}^\pi_{\boldsymbol{\theta}}\big\{\big|\bar{D}_{ik^*} - \bar{D}_{ik}\big| > \eta\big\},
\end{aligned} \tag{C.25}$$

where: (a) follows because $L(k^*) = L(\tau^*_{j+1} - q - 1) \geq \tau^*_j + 1$ on B_{jq}, and (b) follows by the union bound. Note that there are no change-points between cycles $\tau^*_j + 1$ and $k^* = \tau^*_{j+1} - q - 1$. Letting $y^* := \theta_{n(\tau^*_j+1)+1}$, we therefore have

$$\begin{aligned}
\bar{D}_{ik^*} - \bar{D}_{ik} &= \frac{1}{m} \sum_{t \in \mathcal{X}_{ik^*}} D_t \;-\; \frac{1}{m} \sum_{s \in \mathcal{X}_{ik}} D_s \\
&= \frac{1}{m} \sum_{t \in \mathcal{X}_{ik^*}} (\tilde{X}_i \cdot y^* + \epsilon_t) \;-\; \frac{1}{m} \sum_{s \in \mathcal{X}_{ik}} (\tilde{X}_i \cdot y^* + \epsilon_s) \\
&= \bar{\epsilon}_{ik^*} - \bar{\epsilon}_{ik},
\end{aligned} \tag{C.26}$$

for all $k = \tau^*_j + 1, \ldots, k^*$, where: $\tilde{X}_i = \left[\begin{smallmatrix} 1 \\ x_i \end{smallmatrix}\right]$ and $\bar{\epsilon}_{ik} = m^{-1} \sum_{t \in \mathcal{X}_{ik}} \epsilon_t$ for all i, k. Thus, (C.25) implies

$$\begin{aligned}
\mathbb{P}^\pi_{\boldsymbol{\theta}}\big\{\chi_{k^*+1} = 1\,,\, B_{jq}\big\} &\leq \sum_{i=1}^{2} \; \sum_{k=\tau^*_j+1}^{k^*-1} \mathbb{P}^\pi_{\boldsymbol{\theta}}\big\{\big|\bar{\epsilon}_{ik^*} - \bar{\epsilon}_{ik}\big| > \eta\big\} \\
&\leq \sum_{i=1}^{2} \; \sum_{k=\tau^*_j+1}^{k^*-1} \Big(\mathbb{P}^\pi_{\boldsymbol{\theta}}\big\{\big|\bar{\epsilon}_{ik^*}\big| > \tfrac{1}{2}\eta\big\} + \mathbb{P}^\pi_{\boldsymbol{\theta}}\big\{\big|\bar{\epsilon}_{ik}\big| > \tfrac{1}{2}\eta\big\}\Big) \\
&\overset{\text{(c)}}{\leq} 4 \sum_{i=1}^{2} \; \sum_{k=\tau^*_j+1}^{k^*-1} T^{-3/2} \\
&\leq 8T^{-3/2} \left\lfloor \frac{T}{n} \right\rfloor \\
&\leq 8n^{-1}T^{-1/2},
\end{aligned} \tag{C.27}$$

for $k^* = \tau^*_{j+1} - q - 1$, where (c) follows by Lemma 3. Combining (C.24) and (C.27), we get

$$\begin{aligned}\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\{\mathcal{E}_j\} &\le 2+8\sum_{\varepsilon=1}^{\tau^*_{j+1}-\tau^*_j-2}\sum_{q=\varepsilon}^{\tau^*_{j+1}-\tau^*_j-2} n^{-1}T^{-1/2}\\ &\le 2+8n^{-1}T^{-1/2}\left\lfloor\frac{T}{n}\right\rfloor^2\\ &\le 2+8n^{-3}T^{3/2}\\ &\le 2+8\kappa^{-3}. \end{aligned}\tag{C.28}$$

Recalling (C.22), we conclude that

$$\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{\sum_{s=n\hat{\tau}^-_j+1}^{n\tau^*_{j+1}}\left(1-\frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}_{\{s\notin\mathcal{X}\}}\right\} \le (3+8\kappa^{-3})n \le C_2\sqrt{T},\tag{C.29}$$

where $C_2=6\kappa+16\kappa^{-2}$. Q.E.D.

Proof of Lemma 6. Assuming $\tau^*_j<\tau^*_{j+1}-2$, we have

$$\begin{aligned}\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{\sum_{s=n\hat{\tau}^+_j+1}^{n\hat{\tau}^-_j}\left(1-\frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}_{\{s\notin\mathcal{X}\}}\right\} &\overset{(a)}{\le} c_1\sum_{s=n(\tau^*_j+1)+1}^{n\tau^*_{j+1}}\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{\big(\varphi(\theta_s)-p_s\big)^2\,\mathbb{I}_{\{s\notin\mathcal{X},\,n\hat{\tau}^+_j<s\le n\hat{\tau}^-_j\}}\right\}\\ &\le c_2\sum_{s=n(\tau^*_j+1)+1}^{n\tau^*_{j+1}}\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{\|\theta_s-\hat{\theta}_s\|^2\,\mathbb{I}_{\{s\notin\mathcal{X},\,n\hat{\tau}^+_j<s\le n\hat{\tau}^-_j\}}\right\}\\ &= c_2\sum_{s=n(\tau^*_j+1)}^{n\tau^*_{j+1}-1}\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{\|\theta_{s+1}-\hat{\theta}_{s+1}\|^2\,\mathbb{I}_{\{s+1\notin\mathcal{X},\,n\hat{\tau}^+_j\le s<n\hat{\tau}^-_j\}}\right\},\end{aligned}\tag{C.30}$$

where: $c_1=\max_{(\alpha,\beta)\in\Theta}\{4\beta^2/\alpha^2\}$, $c_2=c_1\max_{i=1,2}\big\{\max_\theta\{(\partial\varphi(\theta)/\partial\theta_i)^2\}\big\}$, and (a) follows by definitions of $r(\cdot,\cdot)$, $r^*(\cdot)$, and $\varphi(\cdot)$. Recalling (3.4), we have

$$\hat{\theta}_{s+1}-\theta_{s+1} = \big(\mathcal{J}^s_s\big)^{-1}\sum_{q=1}^{s}w^s_qX_qX_q^{\mathsf{T}}(\theta_q-\theta_{s+1}) + \big(\mathcal{J}^s_s\big)^{-1}\mathcal{M}^s_s \quad \text{for all } s.\tag{C.31}$$

For any given $s=n\hat{\tau}^+_j,\ldots,n\hat{\tau}^-_j-1$ satisfying $s+1\notin\mathcal{X}$, we know that $w^s_q=0$ for $1\le q\le n\hat{\tau}^+_j\le s$, and that $\theta_q-\theta_{s+1}=0$ for $n(\tau^*_j+1)\le q\le s\le n\tau^*_{j+1}$. Because $\tau^*_j<\hat{\tau}^+_j<\hat{\tau}^-_j\le\tau^*_{j+1}$, we deduce that

$$\hat{\theta}_{s+1}-\theta_{s+1} = \big(\mathcal{J}^s_s\big)^{-1}\mathcal{M}^s_s \quad \text{for } s=n\hat{\tau}^+_j,\ldots,n\hat{\tau}^-_j-1 \text{ satisfying } s+1\notin\mathcal{X}.\tag{C.32}$$

Note that $\mathcal{I}^s_s=2\big(\lceil s/n\rceil-L(s/n)+1\big)m=2\big(\lceil s/n\rceil-\hat{\tau}^+_j+1\big)m$ for $s=n\hat{\tau}^+_j,\ldots,n\hat{\tau}^-_j-1$ satisfying $s+1\notin\mathcal{X}$. Hence, (C.32) implies that

$$\begin{aligned}\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\{\|\hat{\theta}_{s+1}-\theta_{s+1}\|^2\,\mathbb{I}_{\{s+1\notin\mathcal{X},\,n\hat{\tau}^+_j\le s<n\hat{\tau}^-_j\}}\} &\le \mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{\big\|\big(\mathcal{J}^s_s\big)^{-1}\mathcal{M}^s_s\big\|^2\,\mathbb{I}_{\{s+1\notin\mathcal{X},\,n\hat{\tau}^+_j\le s<n\hat{\tau}^-_j\}}\right\}\\ &\le \mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{\frac{6}{\rho\big(\lceil s/n\rceil-\hat{\tau}^+_j+1\big)m}\,\mathbb{I}_{\{s+1\notin\mathcal{X},\,n\hat{\tau}^+_j\le s<n\hat{\tau}^-_j\}}\right\},\end{aligned}\tag{C.33}$$

by the arguments used to prove inequality (B.14) and Lemma 2. Summing both sides of (C.33) over $s = n(\tau_j^* + 1), \ldots, n\tau_{j+1}^* - 1$, we deduce that

$$
\begin{aligned}
\sum_{s=n(\tau_j^*+1)}^{n\tau_{j+1}^*-1} \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{ \|\theta_{s+1} - \hat{\theta}_{s+1}\|^2 \, \mathbb{I}_{\{s+1 \notin \mathcal{X},\, n\hat{\tau}_j^+ \le s < n\hat{\tau}_j^-\}} \Big\} & \\
\le \sum_{s=n(\tau_j^*+1)}^{n\tau_{j+1}^*-1} \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{ \frac{6}{\rho\big(\lceil s/n \rceil - \hat{\tau}_j^+ + 1\big)m} \, \mathbb{I}_{\{s+1 \notin \mathcal{X},\, n\hat{\tau}_j^+ \le s < n\hat{\tau}_j^-\}} \Big\} & \\
\le \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Bigg\{ \sum_{s=n\hat{\tau}_j^+}^{n\hat{\tau}_j^- - 1} \frac{6}{\rho\big(\lceil s/n \rceil - \hat{\tau}_j^+ + 1\big)m} \Bigg\} & \\
\overset{\text{(b)}}{\le} n \sum_{q=2}^{\tau_{j+1}^* - \tau_j^* + 1} \frac{6}{\rho q m} & \\
\le \frac{6n}{\rho m} \log(\tau_{j+1}^* - \tau_j^* + 1) & \\
\overset{\text{(c)}}{\le} 6\rho^{-1}\sqrt{T}, & \qquad \text{(C.34)}
\end{aligned}
$$

for $T \ge 3$, where: (b) follows by expressing the time index as $s = (\tau_j^* + q - 1)n + i$ and $\tau_j^* < \hat{\tau}_j^+ < \hat{\tau}_j^- \le \tau_{j+1}^*$, and (c) follows because $m \ge \kappa \log T \ge 2\kappa \log(\tau_{j+1}^* - \tau_j^* + 1)$ for $T \ge 3$, and $n \le 2\kappa\sqrt{T}$. Combining (C.30) and (C.34), we conclude that

$$
\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Bigg\{ \sum_{s=n\hat{\tau}_j^+ + 1}^{n\hat{\tau}_j^-} \bigg(1 - \frac{r(p_s, \theta_s)}{r^*(\theta_s)}\bigg) \mathbb{I}_{\{s \notin \mathcal{X}\}} \Bigg\} \le 6c_2\rho^{-1}\sqrt{T}. \quad \textit{Q.E.D.} \qquad \text{(C.35)}
$$

Proof of Theorem 4. By (4.11) and Lemmas 4, 5, and 6, we have

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Bigg\{ \sum_{t=1}^{T} \bigg(1 - \frac{r(p_t, \theta_t)}{r^*(\theta_t)}\bigg) \mathbb{I}_{\{t \notin \mathcal{X}\}} \Bigg\} &\le (\mathcal{C} + 1)(C_1 + C_2 + C_3)T^{1/2} \\
&\le (\bar{C} + 1)(C_1 + C_2 + C_3)T^{1/2}, \qquad \text{(C.36)}
\end{aligned}
$$

for all $T \ge 3$. Therefore, (4.10) implies $\Delta_{\boldsymbol{\theta}}^{\pi}(T) \le CT^{1/2}\log T$ for all $T \ge 3$, where $C = 8 + (\bar{C} + 1)(C_1 + C_2 + C_3)$. Q.E.D.

Appendix D: Proof of the results in Section 5.

Proof of Theorem 5. In the proof of Theorem 1, let $N = \lceil k_0 T^{2(1-\nu)/3} \rceil$, instead of $N = \lceil k_0 T^{2/3} \rceil$, where $k_0 = 4^{2/3} B^{-2/3}$. Repeating the same arguments from (A.1) to (A.10), deduce that $\sup\{\Delta_{\boldsymbol{\theta}}^{\pi}(T) : V_{\boldsymbol{\theta}}(T) \le BT^{\nu}\} \ge \frac{1}{2} k_2 N^{-1/2} T$ for a certain constant k_2 independent of T, B, and ν. We therefore conclude that $\sup\{\Delta_{\boldsymbol{\theta}}^{\pi}(T) : V_{\boldsymbol{\theta}}(T) \le BT^{\nu}\} \ge cT^{(2+\nu)/3}$ where $c = \frac{1}{8} k_2 B^{1/3}$. Q.E.D.

Proof of Lemma 7. For $M_\nu(\kappa, x_1, x_2)$, using the arguments in the proof of Lemma B.2 to obtain (B.4), we get $\sum_{t=n^2}^{T-1} \|(\mathcal{J}_t^t)^{-1}\mathcal{W}_t^t\|^2 \le 4n^2 V_{\boldsymbol{\theta}}(T)$. Under condition (5.1), this implies that $\sum_{t=n^2}^{T-1} \|(\mathcal{J}_t^t)^{-1}\mathcal{W}_t^t\|^2 \le 4n^2 BT^{\nu} \le 16\kappa^2 BT^{(2+\nu)/3}$. Letting $c_1 = 16\kappa^2 B$, we get (5.4).

For $W_\nu(\mu,\kappa,x_1,x_2)$, consider (B.6) in the proof of Lemma 1, which still holds under condition (5.1):

$$\begin{aligned}\left(\mathcal{J}_t^t\right)^{-1}\mathcal{W}_t^t \;=\; & \sum_{s\in\mathcal{X}_1} \tfrac{w_s^t}{(x_1-x_2)\mathcal{I}_t^t}\begin{bmatrix} -x_2 & -x_1x_2 \\ 1 & x_1\end{bmatrix}(\theta_s-\theta_{t+1}) \\ & + \sum_{s\in\mathcal{X}_2} \tfrac{w_s^t}{(x_1-x_2)\mathcal{I}_t^t}\begin{bmatrix} -x_1 & -x_1x_2 \\ 1 & x_2\end{bmatrix}(\theta_s-\theta_{t+1}). \end{aligned} \tag{D.1}$$

By the arguments used to derive (B.8) and the fact that $w_s^t \le n^{-2}T^{-2\nu}$ for all $s<t-n^2$, we get

$$\begin{aligned}\sum_{t=n^2}^{T-1}\left\|\left(\mathcal{J}_t^t\right)^{-1}\mathcal{W}_t^t\right\|^2 \;\le\; & 8n^{-6}T^{-4\nu}\sum_{t=n^2}^{T-1} t^2\left(\mathcal{I}_t^t\right)^{-2}\max_{1\le s<t-n^2}\left\|\theta_s-\theta_{t+1}\right\|^2 \\ & + 8n^2\sum_{t=n^2}^{T-1}\left(\mathcal{I}_t^t\right)^{-2}\max_{t-n^2\le s\le t}\left\|\theta_s-\theta_{t+1}\right\|^2. \end{aligned}\tag{D.2}$$

Under $W_\nu(\mu,\kappa,x_1,x_2)$, we have $\mathcal{I}_t^t \ge c_\mu n$ for all $t\ge n^2$, where c_μ is a constant independent of T, B, and ν. Hence, the preceding inequality implies that

$$\begin{aligned}\sum_{t=n^2}^{T-1}\left\|\left(\mathcal{J}_t^t\right)^{-1}\mathcal{W}_t^t\right\|^2 \;\le\; & 8c_\mu^{-2}n^{-8}T^{-4\nu}\sum_{t=n^2}^{T-1} t^2\max_{1\le s<t-n^2}\left\|\theta_s-\theta_{t+1}\right\|^2 \\ & + 8c_\mu^{-2}\sum_{t=n^2}^{T-1}\max_{t-n^2\le s\le t}\left\|\theta_s-\theta_{t+1}\right\|^2. \end{aligned}\tag{D.3}$$

Therefore, by condition (5.1) and the fact that $n=\lceil \kappa T^{(1-\nu)/3}\rceil$, the first term on the right hand side of the preceding inequality is bounded above by $8c_\mu^{-2}n^{-8}T^{-4\nu}\sum_{t=n^2}^{T-1} t^2 V_{\boldsymbol{\theta}}(T) \le 8c_\mu^{-2}n^{-8}BT^{3-3\nu} \le 8c_\mu^{-2}\kappa^{-8}BT^{(1-\nu)/3}$. Furthermore, by (B.4) and the fact that $n=\lceil \kappa T^{(1-\nu)/3}\rceil$, the second term is less than or equal to $8c_\mu^{-2}n^2V_{\boldsymbol{\theta}}(T)\le 32c_\mu^{-2}\kappa^2BT^{(2+\nu)/3}$. Thus, the right hand side of (D.3) is bounded above by $8c_\mu^{-2}(\kappa^{-8}+4\kappa^2)BT^{(2+\nu)/3}$. Q.E.D.

Proof of Theorem 6. Note that inequalities (B.12) and (B.15) in the proof of Theorem 2 are valid under condition (5.1), implying that

$$\Delta_{\boldsymbol{\theta}}^{\pi}(T) \;\le\; \frac{3T}{n} \;+\; n^2 \;+\; 2K_0c_2\sum_{t=n^2}^{T-1}\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\|(\mathcal{J}_t^t)^{-1}\mathcal{W}_t^t\|^2 \;+\; \frac{24K_0T}{\rho\tilde{c}n}. \tag{D.4}$$

By Lemma 7, the preceding inequality leads to $\Delta_{\boldsymbol{\theta}}^{\pi}(T) \le 3T/n + n^2 + 2K_0c_1c_2T^{(2+\nu)/3} + 24K_0c_2T/(\rho\tilde{c}n)$. Because $n=\lceil \kappa T^{(1-\nu)/3}\rceil$, this implies $\Delta_{\boldsymbol{\theta}}^{\pi}(T)\le C\,T^{(2+\nu)/3}$ for all $\boldsymbol{\theta}\in\mathcal{V}(T,B)$, where $C=3/\kappa+4\kappa^2+2K_0c_1c_2+24K_0c_2/(\rho\tilde{c}\kappa)$. Q.E.D.

Acknowledgments. The authors thank the Editor-in-Chief Prof. Jim Dai, the Area Editor Prof. Adam Shwartz, the associate editor, and two referees for their helpful comments that improved the presentation and structuring of the paper. Financial support from Duke University Fuqua School of Business is gratefully acknowledged.

References

[1] Araman V, Caldentey R (2009) Dynamic Pricing for Nonperishable Products with Demand Learning. *Operations Research* 57(5):1169–1188.

[2] Aviv Y, Pazgal A (2005) A Partially Observed Markov Decision Process for Dynamic Pricing. *Management Science* 51(9):1400–1416.

[3] Balvers R, Cosimano T (1990) Actively Learning About Demand and the Dynamics of Price Adjustment. *The Economic Journal* 100(402):882–898.

[4] Beck G, Wieland V (2002) Learning and Control in a Changing Economic Environment. *Journal of Economic Dynamics and Control* 26:1359–1377.

[5] Benveniste A, Métivier M, Priouret P (1990) *Stochastic Approximations and Adaptive Algorithms* (Springer-Verlag).

[6] Besbes O, Gur Y, Zeevi A (2015) Non-stationary Stochastic Optimization. Working paper, Columbia University, New York, NY.

[7] Besbes O, Zeevi A (2009) Dynamic Pricing Without Knowing the Demand Function: Risk Bounds and Near-Optimal Algorithms. *Operations Research* 57(6):1407–1420.

[8] Besbes O, Zeevi A (2011) On the Minimax Complexity of Pricing in a Changing Environment. *Operations Research* 59(1):66–79.

[9] Besbes O, Zeevi A (2015) On the (Surprising) Sufficiency of Linear Models for Dynamic Pricing with Demand Learning. *Management Science* 61(4):723–739.

[10] Broder J, Rusmevichientong P (2012) Dynamic Pricing under a General Parametric Choice Model. *Operations Research* 60(4):965–980.

[11] Brown R (1956) Exponential Smoothing for Predicting Demand. Presented at the Tenth National Meeting of the Operations Research Society of America, San Francisco, November 16, 1956.

[12] Chen Y, Farias V (2013) Simple Policies for Dynamic Pricing with Imperfect Forecasts. *Operations Research* 61(3):612–624.

[13] den Boer A (2015) Tracking the Market: Dynamic Pricing and Learning in a Changing Environment. *European Journal of Operational Research* 247(3):914–927.

[14] den Boer A, Zwart B (2014*a*) Simultaneously Learning and Optimizing using Controlled Variance Pricing. *Management Science* 60(3):770–783.

[15] den Boer A, Zwart B (2014*b*) Mean Square Convergence Rates for Maximum Quasi-likelihood Estimators. *Stochastic Systems* 4(2):375–403.

[16] Farias V, van Roy B (2010) Dynamic Pricing with a Prior on Market Response. *Operations Research* 58(1):16–29.

[17] Garivier A, Moulines E (2011) On Upper-Confidence Bound Policies for Switching Bandit Problems. *Proceedings of the 22nd International Conference on Algorithmic Learning Theory (ALT)* 174–188.

[18] Harrison J, Keskin N, Zeevi A (2012) Bayesian Dynamic Pricing Policies: Learning and Earning Under a Binary Prior Distribution. *Management Science* 58(3):570–586.

[19] Harrison J, Sunar N (2015) Investment Timing with Incomplete Information and Multiple Means of Learning. *Operations Research* 63(2):442–457.

[20] Holt C (1957) Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages. *Office of Naval Research Memorandum* 52.

[21] Keller G, Rady S (1999) Optimal Experimentation in a Changing Environment. *The Review of Economic Studies* 66(3):475–507.

[22] Keskin N, Zeevi A (2014) Dynamic Pricing with an Unknown Demand Model: Asymptotically Optimal Semi-myopic Policies. *Operations Research* 62(5):1142–1167.

[23] Lai T (1995) Sequential Changepoint Detection in Quality Control and Dynamical Systems. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(4):613–658.

[24] Lai T (2003) Stochastic Approximation. *The Annals of Statistics* 31(2):391–406.

[25] Lobo M, Boyd S (2003) Pricing and Learning with Uncertain Demand. Working Paper. Stanford University, Stanford, CA.

[26] Phillips R (2005) *Pricing and Revenue Optimization* (Stanford University Press, Stanford, CA).

[27] Rusmevichientong P, Tsitsiklis J (2010) Linearly Parameterized Bandits. *Mathematics of Operations Research* 35(2):395–411.

[28] Rustichini A, Wolinsky A (1995) Learning About Variable Demand in the Long Run. *Journal of Economic Dynamics and Control* 19:1283–1292.

[29] Shiryaev A (2010) Quickest Detection Problems: Fifty Years Later. *Sequential Analysis* 29(4):345–385.

[30] Tsybakov A (2009) *Introduction to Nonparametric Estimation* (Springer, New York).

[31] Wang Z, Deng S, Ye Y (2014) Close the Gaps: A Learning-while-doing Algorithm for a Class of Single-product Revenue Management Problems. *Operations Research* 62(2):318–331.

[32] Winters P (1960) Forecasting Sales by Exponentially Weighted Moving Averages. *Management Science* 6(3):324–342.